

Danish University Colleges

Chapter 11: Evaluation regarding implementation of the HAND in HAND programme

Nielsen, Birgitte Lund

Published in:
The HAND in HAND Programme Evaluation Report

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):
Nielsen, B. L. (2020). Chapter 11: Evaluation regarding implementation of the HAND in HAND programme. In S. Vieluf, M. Rozman, & N. Roczen (Eds.), *The HAND in HAND Programme Evaluation Report* (pp. 268-283).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Download policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The HAND in HAND Programme

Evaluation Report

Svenja Vieluf, Mojca Rožman & Nina Roczen



Contents

- Foreword 3
- PART A: INTRODUCTION 4
 - Chapter 1: The HAND in HAND Project and its aims 5
 - Chapter 2: The evaluation of the HAND in HAND programme 17
 - Chapter 3: Development of the assessment for use in evaluation of the HAND in HAND programme 24
 - Chapter 4: The HAND in HAND Field Trials: samples and research questions 46
- PART B: RESULTS OF THE SUMMATIVE EVALUATION 58
 - Chapter 5: Change in social and emotional competencies and intercultural competencies/diversity awareness: Results from questionnaires 59
 - Chapter 6: Changes in intercultural competence/diversity awareness: Results from a vignette study 136
 - Chapter 7: Effects of the HAND in HAND programmes on classroom and school climates 189
 - Chapter 8: Participants’ summative evaluation of the HAND in HAND programmes 215
- PART C: RESULTS OF FORMATIVE EVALUATION 231
 - Chapter 9: Participants’ ideas for improving the HAND in HAND programmes 232
 - Chapter 10: Quality assurance in the HAND in HAND project 255
 - Chapter 11: Evaluation regarding implementation of the HAND in HAND programme 268
- PART D: SUMMARY AND INTEGRATION OF FINDINGS 284
 - Chapter 12: How effective are the HAND in HAND Programmes and how could they be improved for future use? 285

Foreword

This report presents the results of an evaluation of the HAND in HAND programme. The HAND in HAND programme aimed at building more inclusive schools by fostering the social, emotional and intercultural (SEI) competencies of students and school staff. The evaluation had the aim to determine how effective the programme was in achieving this aim, whether it had unintended effects, what participants themselves thought about the programme, and what we learned in the evaluation process about possible starting points for a further improvement of the programme. The evaluation report is structured into four sections: First, an introduction. Second, a section presenting the results of the summative outcome evaluation. Third, a section presenting the results of formative evaluations. Fourth, a summary of findings and conclusions concerning the quality of the HAND in HAND programme as well as suggestions for changes. Each of these section is organised into several sub-sections. The introductory chapter starts with a description of the HAND in HAND project and its aims. In the second chapter our approach to evaluating the HAND in HAND programme is set out. In a third chapter the development of the assessment for use in the evaluation is described. The samples and research questions for the HAND in HAND field trials are topic of the fourth chapter. Consequently evaluation results are presented. The second section starts with two chapters that present analyses of changes in social and emotional competencies and intercultural competencies/diversity awareness based on self-report measures and based on vignettes respectively. These are followed by a chapter on effects of the HAND in HAND programmes on the quality of classroom climates. A fourth chapter in this section focuses on participants' view on the quality of the programme and present results from semi-structured focus-group interviews. The third section presents formative evaluation components. It includes, first, a chapter that summarizes participants' ideas for improving the HAND in HAND programmes. Second, a chapter describing the quality assurance procedures implemented during the HAND in HAND project. The third chapter deals with the quality of the implementation of the HAND in HAND programmes. The evaluation report ends with a summary of results that aims at answering two broad questions: Did the programmes have the intended effects? And: How could the programmes be improved?

Svenja Vieluf, Mojca Rožman and Nina Roczen

PART A:

INTRODUCTION

Chapter 1:
The HAND in HAND Project and its aims¹

Ana Kozina, Maša Vidmar, Manja Veldin

¹ This text is a part of the publication *Social, emotional and intercultural competencies for inclusive school environments across Europe* (Kozina, 2020) where a longer text with more information on core concepts and the project itself can be found.

1.1. Aims of the HAND in HAND project

The HAND in HAND project targeted the need detected in Europe and internationally to develop inclusive societies (schools and classrooms) that allow every student to feel accepted and be able to achieve their potential, particularly in response to increasing migration flows. HAND in HAND seeks to achieve this by fostering the social, emotional and intercultural (SEI) competencies of students and school staff – the whole-school approach. The whole school approach engages the entire school community (in our case, the students of a single class, their teachers, school counsellors, and the principal) as part of a cohesive, collective and collaborative effort. The project aimed to pilot a programme, to help develop these competencies and propose a system-level solution for upscaling at the national and European levels. Accordingly, the consortium has developed an open-access systemic policy tool: EU-based, universal SEI learning programmes (HAND in HAND programmes: a HAND in HAND programme for students (Marušić et al, 2019) and a separate HAND in HAND programme for school staff (Jensen & Gøtzsche, 2020).

1.2. Scientific background of the HAND in HAND project

Regarding the social and emotional competencies, the work of the USA-based Collaborative for Academic, Social and Emotional Learning (CASEL, 2013) served us as a foundation although bringing a more humanistic (relational) perspective, and for the intercultural competencies/diversity awareness the work of several authors acted as a scientific background (Bennett, 1986, 1993, 2004, 2014; Blell & Doff, 2014; Byram, 1997; Deardorff, 2006). Building on previous theories, the final core concepts and definitions have been agreed following extensive discussions based on the expertise held by the project team and a literature review that had been performed at the beginning of the project.

1.2.1. Social, emotional and intercultural competencies

Social and emotional competencies and intercultural competencies are usually treated separately within various research traditions, although they considerably overlap (Nielsen et al., 2019). Even though the social and emotional components are often included in the core of intercultural competencies/diversity awareness (e.g. Stier, 2003), there is only a small overlap in research. In HAND in HAND, we place a strong focus on the constructs important for both areas (e.g. openness, respect, relations) while also focusing on parts that are more specific to each (e.g. self-awareness in the social and emotional part and moving beyond the self–other binary in the intercultural part).

1.2.2. Social and emotional competencies

CASEL (2013) outlines five interrelated dimensions (clusters of competencies) of social/emotional competencies held by students that have also been applied to school staff (Schonert-Reichl, Hanson-Peterson et al., 2015): self-awareness; self-management; social awareness; relationship skills; and responsible decision-making. In addition to CASEL's dimensions and intercultural competencies/diversity awareness, another dimension was included for school staff. Given the strong relational orientation of the core HAND in HAND concepts and the programme, it was needed to include an additional SEI dimension for school staff – relational competence. This competence overlaps with several SEI dimensions and is much broader than CASEL's relationship skills; it also brings a humanistic orientation concentrating on the importance of the student-teacher relationship and what happens within that relationship (see below) and was thus conceptualised as a separate entity.

Following the CASEL Guide (2013), **self-awareness** is the ability to recognise one's emotions and thoughts and their influence on behaviour. This includes accurately assessing one's strengths and limitations and possessing a well-grounded sense of confidence and optimism. In the updated framework (Weissberg, Durlak, Domitrovich & Gullotta, 2015), the ability to understand one's own personal goals and values, and having a positive mind-set is added. In HAND in HAND, we have reflected on self-awareness as the ability to recognise one's emotions, bodily sensations and thoughts and their influence on how we respond. This includes having a sober, accepting/recognising way of looking at oneself; and the will and continuing wish to work on establishing all of it. Self-awareness is reflected in being present in your body, thoughts and feelings in a non-judgmental manner, e.g. being mindful. In HAND in HAND's conceptualisations, we also see it as not so much a goal and an outcome as an ongoing process that continues to happen (not something that is achieved or completed and is then 'available for further use').

Self-management is the ability to regulate one's emotions, thoughts and behaviours effectively in different situations. This includes managing stress, controlling impulses, motivating oneself, and setting and working toward achieving personal and academic goals (CASEL, 2013). The updated CASEL framework (Weissberg et al, 2015) includes the ability to delay gratification and perseverance through challenges. In HAND in HAND, we understand self-management as the ability to regulate one's emotions, bodily sensations, and thoughts and their influence on how we react.

Social awareness is the ability to adopt the perspective of and empathise with others from diverse backgrounds and cultures, to understand social and ethical norms of behaviour, and to recognise family, school and community resources and supports (CASEL, 2013). The updated framework (Weissberg et al., 2015) also includes compassion. In HAND in HAND's conceptualisations, social awareness is the ability to take on the perspective of and to have empathy and compassion for others from diverse backgrounds and cultures, to understand, accept and recognise social and ethical norms of behaviour, to be aware of cultural synergies overcoming the self/other binary and making space for different points of view, also recognising the influence and importance of family, school and community. In the part "recognising the influence and importance of family, school and community", we wish to stress that this influence is not always supportive, although we still need to recognise the contextual factors. As such, it also holds strong intercultural/transcultural momentum by incorporating the perspective of others, not only to understand but also to accept and recognise it, along with the importance of making space for the differences between perspectives.

Relationships skills are the ability to establish and maintain healthy and rewarding relationships with various individuals and groups. This includes communicating clearly, listening actively, cooperating, resisting inappropriate social pressure, negotiating conflict constructively, and seeking and offering help when needed (CASEL, 2013). The updated framework (Weissberg et al., 2015) also includes acting according to social norms. In HAND in HAND's conceptualisations, relationship skills are the ability to establish and maintain constructive relationships and the will to persist, even when it seems impossible to maintain them. It is important to stress the will to persist because these skills are especially challenged and needed in difficult times. This includes the ability to accept personal and social responsibility and go into the relationship with personal presence, aware that in a constructive relationship, individual needs to establish synergy between taking care of their integrity and taking care of the group (Juul & Jensen, 2010).

Responsible decision-making is the ability to make constructive and respectful choices about personal behaviour and social interactions based on a consideration of ethical standards, safety concerns, social norms, a realistic evaluation of the consequences of various actions, and the well-being of self and others (CASEL, 2013). In HAND in HAND, we add to that the importance of knowledge of social groups and their products and practices beyond self/other, and knowledge about asymmetrical and global cultural processes (e.g. unequal positions). Once again, we can see the intercultural/transcultural aspect being added.

Intercultural competencies²: as we have seen, intercultural competencies and social/emotional competencies are related although thus far there has not been a specific intercultural/transcultural focus in social and emotional learning research (for a review, see Nielsen et al., 2019). Social, emotional competencies play a central role in various models of intercultural competencies (e.g. Deardoff, 2006). Based on the literature review, we included models that are well-elaborated, internationally recognised, general, i.e. not limited to only one field, offer clearly defined concepts and/or outcomes, take a developmental perspective and have empirical support. Thus, HAND in HAND's conceptualisation of intercultural competencies brings together the PISA model of global competence (OECD, 2018), Deardorff's model (Deardoff, 2006), Byram's model of intercultural communicative competence (Byram, 1997) as well as Bennett's developmental model of intercultural sensitivity (Bennett, 1986, 1993, 2004, 2014). In a broader sense, intercultural competencies/diversity awareness are defined as the ability to communicate effectively and appropriately in intercultural situations, based on one's: intercultural knowledge (e.g. self-awareness, understanding and knowledge of intersectional differences); competencies (e.g. seeing from others' perspectives; listening, observing and interpreting; analysing, evaluating and relating; ability to interpret a document or event arising from various cultures; ability to acquire new knowledge concerning a culture and culture practices), and attitudes (respect – valuing cultural diversity; openness – to intercultural learning and people from diverse cultural backgrounds; withholding judgement; curiosity and discovery – tolerating ambiguity and uncertainty). In addition, we took into account Blell & Doff 's Model of Inter- and Transcultural Communicative Competence (I/TCC) (Blell & Doff, 2014). This model is built on traditional models of intercultural communication competence (Byram, 1997). However, it suggests moving beyond a self-other binary to an understanding of culture and cultural identity as being hybrid, dynamic and multifaceted (e.g. having and recognising multiple flexible identities of one self and others, multiple ways in which they are expressed and how these influence us being together).

1.2.3. Relational competence

Alongside Social and emotional competencies and intercultural competencies/diversity awareness in the HAND in HAND programme for school staff, relational competence was used as a core feature. Relational competence is promoted by the development of Social and emotional competencies and

² Please note that throughout the rest of the report this domain is referred to as „intercultural competence/diversity awareness“. Why this is the case, is described in chapter 3.

intercultural competencies/ diversity awareness and at the same time is specific to professionals (e.g. teachers, counsellors, principals). It is defined as a professional's ability to 'see' the individual child on its own terms and attune their behaviour accordingly, without giving up leadership, as well as the ability to be authentic in the contact with the child. It is also crucial that professionals have the ability and will to take full responsibility for the quality of the relationship (Juul & Jensen, 2017). The relational competence held by teachers is regarded as the foundation for creating an inclusive environment in the classroom that enables the Social and emotional competencies and intercultural competencies/ diversity awareness of both students and teachers to be developed (Jensen, Skibsted, & Christensen 2015; Juul & Jensen 2017).

1.2.4. The whole-school approach

The whole-school approach engages both students and the school staff in the building of an inclusive and supportive environment by directly influencing the quality of the relationship between students and teachers via the promotion of their Social and emotional competencies and intercultural competencies/ diversity awareness. The importance of relationships is reflected in the concept of the classroom climate. Classroom climate refers to the shared perception held by students and teachers concerning the quality of the classroom learning environment (Adelman & Taylor, 2005; Fraser, 1989) and has three main components (Moos, 1979): (i) *Relationship*: the quality of personal relationships (between teachers and students, as well as between students) within the environment: the extent to which people are involved in the environment and support/help each other and treat each other with respect; (ii) *Personal development*: the extent to which an environment is in place that supports the personal growth and self-enhancement of each individual in this environment; (iii) *System maintenance and change*: the extent to which the environment is orderly, clear with respect to expectations, maintains control, and is responsive to change. According to offer-take-up models of teaching (Fend, 1998; Helmke, 2006), classroom climate is the outcome of the complex interplay of teacher behaviours (the learning offer) and student behaviours (their take-up of such offers) that are both influenced by individual characteristics of all actors, characteristics of the school's broader context, the neighbourhood, the school system, and by situational and interactional factors.

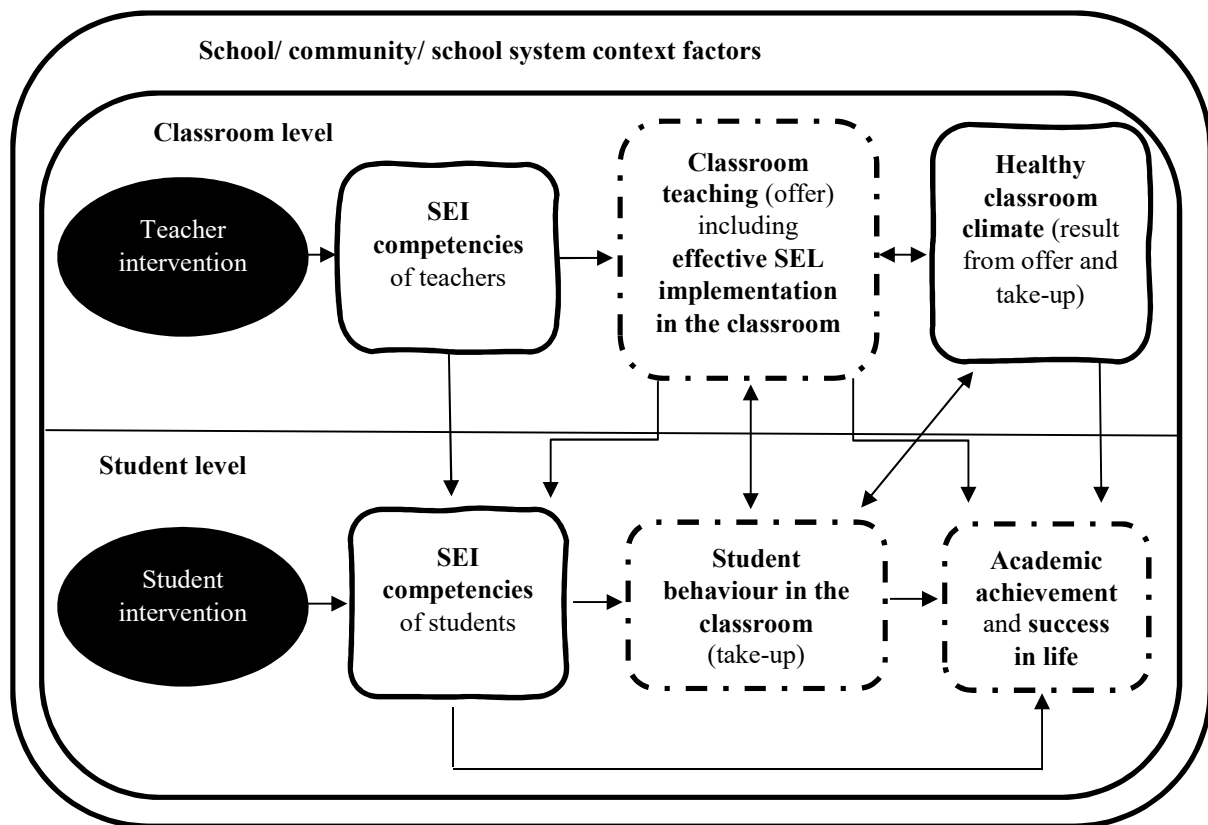


Figure 1.1. The whole-school approach used in the HAND in HAND

The whole-school approach as understood in HAND in HAND is based on the Prosocial Classroom model (Jennings & Greenberg, 2009) combined with offer-take-up models of teaching (Fend, 1998; Helmke, 2006). The Prosocial classroom model explains the link between teacher social/emotional competencies and outcomes at the classroom and student levels. Teachers' social and emotional competencies impact students in at least three ways: (1) teacher's competencies influence the quality of the teacher-student relationship, (2) the teacher serves as a role model of social/emotional competencies for students; and (3) the teacher's social/emotional competencies influence management of the classroom. Together, these factors co-create a healthy classroom climate that fosters students' social, emotional and learning achievement. The model also explains how teachers' social/emotional competencies are important for their well-being. A teacher with developed social/emotional competencies (e.g. one capable of high self-awareness and self-management) is able to manage their daily social/emotional challenges (e.g. inappropriate, abusive student behaviour, non-participation, troubled parents, etc.) that arise in their work, making teaching easier and the teacher feel more effective in their role. But the opposite can also happen; teachers' poor social/emotional competencies lead to poor student relationships and classroom management problems. This can produce a negative climate that prevents the achievement of educational and developmental goals. As

a result, the teacher may experience a sense of inefficiency and emotional exhaustion, in turn weakening their daily social and emotional capacities and further degrading classroom relationships and the quality of leadership, the climate, and the achievement of goals (creating a ‘burnout cascade’). The models also show the context in which the teacher performs (class or subject level, leadership support, school safety, involvement in the local community, etc.) is also important.

1.3. The HAND in HAND programmes

The HAND in HAND programmes consist of three interconnected programmes: a HAND in HAND programme for students and a separate HAND in HAND programme for school staff. The HAND in HAND student programme (Marušić et al., 2020) is organised in five modules, each lasting 90 minutes. Each module focuses on one of the core socio-emotional competencies according to CASEL (2003) and includes an exercise aimed at developing intercultural competencies/diversity awareness. More details can be found in Jugović et al. (2020). The HAND in HAND programme for school staff consists of a programme for teachers and a separate programme for school leaders and counsellors (Jensen et al., 2018a; Jensen et al., 2018b). The programme for teachers has four modules: two modules lasting 2 days and another two modules each lasting 1 day. The programme for the school leaders and counsellors requires 2 single days. More details can be found in Jensen and Gøtzsche (2020).

In the HAND in HAND programmes short theoretical inputs alternated with practical exercises. These exercises fell into four categories:

- (1) inner exercises, in particular: body scans, which were led by the trainer and practised in the whole group;
- (2) physical exercises and games, e.g. counting up to 20 in a group, shaking arms and legs, dancing, passing a ball from head to head, climbing up and down on a chair, balancing on one’s toes, or giving each other massage; these exercises were also led by the trainer and practised in the whole group or in pairs;
- (3) exercises with discussions or dialogues, e.g. discussions about how to recognise emotions, listening to another’s story and trying to reproduce it without commenting, telling a story together by taking turns and each time taking up what the other had said, practising the formulation of “I”-

messages, or structure dialogues to reflect one's own pedagogical practice (only the teachers); these exercises were often done in pairs or small groups;

- (4) exercises addressing diversity, e.g. experiential exercises where students experienced in games how it felt while entering a group without knowing the rules according to which the group was behaving, or how it felt while they were treated on the basis of prejudices about a social difference category, or how it felt when they lacked privileges that all other children had. This category also includes teacher reflections on their own way of addressing diversity in schools. Reflection on diversity was done in pairs, small groups or in the whole group.

All these exercises were led by the trainers. At the end of each exercises the trainers reflected with the participants on experiences during the exercises.

1.4. References

- Adelman, H. S., & Taylor, L. (2005). Classroom climate. In S. W. Lee (Ed.), *Encyclopedia of school psychology*, Thousand Oaks, CA: Sage.
- Bennett, M. J. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations*, 10, 179–95.
- Bennett, M. J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In R. M. Paige (Ed.), *Education for the intercultural experience* (21–71). Yarmouth, ME: Intercultural Press.
- Bennett, M. J. (2004). Becoming interculturally competent. In J. Wurzel (Ed.), *Toward multiculturalism: A reader in multicultural education* (2nd ed., pp. 62–77). Newton, MA: Intercultural Resource.
- Bennett, M. J. (2014). *Intercultural communication*. Sage Publications, Inc. Bidwell.
- Blell, G. & Doff, S. (2014). It takes more than two for this tango: Moving beyond the self/other-binary in teaching about culture in the global EFL-classroom. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 19, 77–96.

- Byram, M. (1997). *Teaching and Assessing Intercultural Communicative Competence*. Clevedon: Multilingual Matters.
- Collaborative for Academic, Social and Emotional Learning (2013). *The 2013 CASEL Guide: Effective social and emotional learning programmes-preschool and elementary school edition*. Chicago, IL: Author.
- Deardorff, D. K. (2006). Identification and Assessment of Intercultural Competence as a Student Outcome of Internationalization. *Journal of Studies in International Education*, 10, 241–266.
- Fend, H. (1998). *Qualität im Bildungswesen*. Weinheim.
- Fraser, B. J. (1989). Twenty years of classroom climate work: Progress and prospect. *Journal of Curriculum Studies*, 21(4), 307–327.
- Helmke, A. (2006). Unterrichtsqualität. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (812–820). Weinheim: Beltz Psychologie Verlags Union.
- Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, 79, 491–525.
- Jensen, E., Skibsted, E. B., & Christensen, M. V. (2015). Educating teachers focusing on the development of reflective and relational competences. *Educational research for policy and practice*, 14(3), 201–212.
- Jensen, H. (2017). Friendship, empathy and mindfulness in children's groups: developing children's natural capacities. In S. Hart, *Inclusion, play and empathy: Neuroaffective development in children's groups* (281–301). London: Jessica Kingsley Publishers.
- Juul, J., & Jensen, H. (2010). *Od poslušnosti do odgovornosti*. Ljubljana: Didakta.
- Jensen, H., Gøtzsche, K., Andersen Réol, L., Dyrborg Laursen, H., Nielsen, B. L., Denk, A., Kozina, A., Vršnik Perše, T., Marušić, I., Jugović, I., Rasmusson, M., & Oskarsson, M. (20120). HAND in HAND programme for school leaders and counsellors, manual. Retrieved from www.handinhand.si
- Jensen, H. & Gøtzsche, K. (2020). Development of the social, emotional and intercultural learning programme for school staff. In A. Kozina (Ed), *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments*

- across Europe. Relationships matter* (83–107). Hamburg: Verlag Dr. Kovač.
- Jugović, I., Puzić, S & Mornar, M. (2020). Developing the social, emotional and intercultural learning programme for students . In A. Kozina (Ed), *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe. Relationships matter* (59–83). Hamburg: Verlag Dr. Kovač.
- Juul, J., & Jensen, H. (2017). *Relational Competence. Towards a New Culture of Education*. Windberg: Mathias Voelchert GmbH Verlag.
- Kozina, A. (Ed) (2020). *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe: Relationships matter*. Hamburg, Germany: Verlag Dr. Kovač (in print).
- Marušić, I., Puzić, S., Jugović, I., Košutić, I., Matić, J., Mornar, M., Nielsen, B. L., Jensen, H., Rasmusson, M., Oskarsson, M., Denk, A., Kozina, A. & Veldin, M. (2020). HAND in HAND programme for students, manual. Retrieved from: <http://handinhand.si>
- Moos, R. H. (1979). *Evaluating educational environments*. San Francisco: Jossey-Bass.
- Nielsen, B. L., Laursen, H. D., Reol, L. A., Jensen, H., Kozina, A., Vidmar, M., ... & Jurko, S. (2019). Social, emotional and intercultural competencies: a literature review with a particular focus on the school staff. *European journal of teacher education*, 42(3), 410–428.
- OECD (2018). *PISA 2018 Global Competence*. Paris: OECD Publishing.
- Roczen, N., Endale, W., Vieluf, S., & Rožman, M. (2020). Development of the assessment for use in evaluation of the HAND in HAND programme. In A. Kozina, *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe* (131-156). Hamburg: Verlag Dr. Kovač
- Schonert-Reichl, K. A., Hanson-Peterson, J. L., & Hymel, S. (2015). Social and emotional learning and pre-service teacher education. In J. A. Durlak, R. P. Weissberg, C. E. Domitrovich, and T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (406–421). NY: Guilford.

Stier, J. (2003). Internationalisation, Ethnic Diversity and the Acquisition of Intercultural Competencies. *Intercultural Education*, 14(1), 77–91.

Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present, and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds), *Handbook of social and emotional learning Research and practice* (3–20). New York, London: The Guilford Press.

Chapter 2: The evaluation of the HAND in HAND programme³

Nina Roczen, Svenja Vieluf, Mojca Rožman

³ Please note that part of the chapter has already been published in Vieluf, Denk, Rožman and Roczen (2020).

2.1. Evaluation: Definition and Functions

The evaluation of the HAND in HAND programme was multifaceted and pursued different objectives. In general, evaluation is defined as “a form of ‘disciplined inquiry’ (Lincoln & Guba, 1986; p. 550) that applies scientific procedures to the collection and analysis of information about the content, structure and outcomes of programmes, projects and planned interventions” (Clarke & Dawson, 1999; p. 1). Evaluations have different functions. On the one hand, they usually aim to determine the “merit or worth” of something (e.g., Scriven, 1967). On the other hand, they are also intended to help “people make wise decisions and choices about future programming” (Weiss, as cited by Clarke & Dawson, 1999; p. 2). While the first function is referred to as “summative”, the second is called “formative” (Black & Wiliam 2003; Wiliam & Thompson 2008). Another distinction frequently made in the literature concerning evaluation is that between the evaluation of processes and outcomes (e.g. Chen, 1996; see Figure 1.1.).

		Evaluation Function	
		<i>Formative</i> (→improvement)	<i>Summative</i> (→assessment)
Program me Stages	<i>Process</i>	Formative process evaluation	Summative process evaluation
	<i>Outcome</i>	Formative outcome evaluation	Summative outcome evaluation

Figure 1. Basic Types of Evaluation

During the 20th century, a strong focus was given to summative outcome evaluations with experimental designs, in the framework of which it is analysed whether an intervention had causal effects on predefined outcomes (see Widmer, 2012). Using randomized control-group experiments can be considered a gold standard for making causal conclusions. However, this strategy to summative outcome evaluation has also been criticised for its one-sidedness, the neglect of processes and the distance to the participants (e.g. Abma, 2006; Greene, 1988; 2001; Guba & Lincoln, 1989; Stake, 1975). For example, Stake (1975) emphasized, in his outline of a responsive evaluation, the importance of taking account of the participants’ perspectives to obtain a deeper understanding of an intervention’s effects. Therefore, the evaluation of the HAND in HAND

programme combined a randomized-control group experiment that allows for assessing causal effects of the programme with semi-structured focus group interviews with participants to allow for understanding the perspectives and experiences of those people the programmes are ultimately intended to serve.

A further theoretical distinction of evaluation types concerns the role of the evaluators with regard to a project. Depending on whether the evaluation is performed by persons directly involved in a programme (e.g., programme developers, trainers) or by persons whose only relation to a programme is to evaluate it, an evaluation can be defined as internal or external, respectively. Internal evaluations have the advantage that the evaluators usually know the context and the internal processes well, an external evaluation is usually attributed with a particularly high degree of objectivity (see Conley-Tyler, 2005). Both internal and external evaluations can further be formative and summative and can relate to both processes and outcomes. The HAND in HAND project was both internally and externally evaluated. The internal evaluation was focused on the processes (project management and programme implementation), whereas the external evaluation mainly focused on the outcomes (the Hand in Hand programmes for students, teachers and school leaders/ other school staff). However, the external evaluation team was also involved in an ongoing process of consultation and negotiation among all project partners and, as part of this process, also provided suggestions with regard to the definition of the aims of the programme (“core constructs”) and with regard to the development of the programme. The other way around, the developers of the programme also contributed to the development of the assessment for the summative and formative outcome evaluation. For this reason, the originally merely external evaluation increasingly became an internal evaluation as the project progressed.

Hence, the evaluation of the Hand in Hand project was a complex process that involved different stakeholders, different perspectives and different methods and followed different aims. Thereby, more comprehensive information about the quality of the project and possible approaches for improving it could be collected as compared to traditional evaluations that focused only on effectiveness. Most importantly, those whom the project is intended to serve, the students, teachers, school leaders and other school staff, also got a voice and a chance to present their perspectives on the programme. This approach has the additional advantage that different levels of programme effects are taken into account. More

specifically, it can be assumed that the success of interventions is generally gradual: In a first stage, satisfaction and experienced relevance of a training can be achieved. This provides the basis for building knowledge and changing convictions and motivation in a second stage. Only at a further stage, building on the latter and depending on other factors, can changes in behaviour take place (cf. Guskey, 2000). Also for those reasons, in the evaluation of the HAND in HAND programme, different strategies were combined.

2.2. The External Evaluation of the HAND in HAND Programme

For the external evaluation, both an experimental outcome evaluation and an interview-based evaluation, a summative and a formative approach, along with quantitative and qualitative data analysis were realized. A randomised control group experiment with pre-post and follow-up measurements had the aim to find out whether the HAND in HAND programme had actually served the purpose it was developed for: fostering the social and emotional competencies and intercultural competencies/ diversity awareness of students, teachers and other school staff and, mediated through this improvement, to improve classroom climates in the participating schools. The results regarding the change in Social and emotional competencies and intercultural competencies/ diversity awareness are presented in Chapter 5 (see also Rožman, Roczen & Vieluf, 2020). Results regarding the change in classroom climate are summarized in Chapter 12 in this book. Complementing this part of the evaluation, semi-structured focus group interviews with groups of all participants (students, teachers and school leaders together with other school staff) inform how participants evaluated the programme, which criteria are relevant in their judgement and how these relate to the criteria predefined by the researchers (see Chapter 12). The semi-structured focus group interviews also give a basis for a formative outcome evaluation of the HAND in HAND programme; namely, for identifying ways to improve it. The respective results are presented in Chapter 12.

2.3. The internal evaluation of the HAND in HAND programme

The internal evaluation had two main focuses. One was an internal evaluation of the implementation of HAND in HAND programmes, which looked at the perspective of the trainers. The trainers filled out reflection logs after each training session, online surveys with open ended questions and a few Likert-type items. Results regarding the challenges the trainers have encountered and the

developments that have taken place are described in Chapter 12 of this book (Nielsen, this publication; see also Nielsen, 2020). The second focus was internal quality assurance. Within this framework, both the general project management and the implementation of the programme at the schools were continuously monitored. Various instruments were used, such as a web tool to keep track of the project progress, quality visits, and questionnaires for assessing the quality of project meetings as well as a continuous dialogue with the project coordination. The results are summarized in Chapter 12 in this book (Rasmusson, Oskarsson, Eliasson, & Dahlström, this publication; see also Rasmusson, Oskarsson, Eliasson, & Dahlström, 2020).

2.4. Conclusions

The evaluation of HAND in HAND programmes was carried out by partners involved in programme development and implementation as well as by external partners and it was both, summative and formative. The focus of the external summative outcome evaluation was on answering the question whether the HAND in HAND programme had effects on desired outcomes (*summative outcome evaluation*). Additionally, the external summative outcome evaluation aimed at understanding the perspectives of participants on the quality of the programmes. The external *formative outcome evaluation* aimed at identifying possibilities to improve the programme from the participants' point of view. For purposes of an internal summative and formative evaluation, the implementation of the programme (*summative process evaluation*) and the quality of the overall project management (*formative process evaluation*) were observed through various surveys and quality visits. This multifaceted evaluation strategy ensures that by assessing different levels of possible programme success and by viewing processes from the perspective of different actors a balanced and comprehensive evaluation of HAND in HAND programmes is achieved.

2.5. References

- Abma, T. A. (2006). The practice and politics of responsive evaluation. *American Journal of Evaluation*, 27(1), 31–43.
- Black, P. & Wiliam, D. (2003). In praise of educational research: formative assessment. *British Educational Research Journal*, 29(5), 623–637.

- Chen, H. T. (1996). A comprehensive typology for programme evaluation. *Evaluation Practice*, 17(2), 121–130.
- Clarke, A., & Dawson, R. (1999). *Evaluation research. An introduction to principles, methods, and practice*. London: Sage.
- Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation?. *Evaluation Journal of Australasia*, 4(1-2), 3-11.
- Denzin, N. K. (2001). *Interpretative interactionism* (2nd ed.). Thousand Oaks, CA: Sage.
- Greene, J. C. (1988). Stakeholder participation and utilization programme evaluation. *Evaluation Review*, 12(2), 91–116.
- Greene, J. C. (2001). Dialogue in evaluation: A relational perspective. *Evaluation*, 7(2), 181-203.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Guskey, T. R. (2000). *Evaluating Professional Development*. Thousand Oaks, CA: Corwin Press.
- Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. In D. D. Williams (Ed.), *Naturalistic evaluation* (pp. 73–84). San Francisco, CA: Jossey-Bass.
- Nielsen, B.L. (2020). Implementing the HAND in HAND programme for school staff. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe* (107-129). Hamburg: Verlag Dr. Kovač.
- Rasmusson, M., Oskarsson, M., Eliasson, N., Dahlström, H. (2020). Quality assurance in the HAND in HAND project. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe* (219-233). Hamburg: Verlag Dr. Kovač.
- Rožman, M., Roczen, N., & Vieluf, S. (2020). Evaluation of the HAND in HAND programme: Results from questionnaire scales. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe* (157-194). Hamburg: Verlag Dr. Kovač.

- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Chicago, IL: Rand McNally.
- Stake, R.E. (1975). To evaluate an arts programme. In R. E. Stake, *Evaluating the arts in education: A responsive approach* (pp. 13-31). Columbus, OH: Merrill.
- Vieluf, S., Denk, A., Rožman, M., & Roczen, N. (2020). How do the participants evaluate the HAND in HAND programme? Results of semi-structured focus group interviews. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (195-218). Hamburg: Verlag Dr. Kovač.
- Widmer, T. (2012). Unabhängigkeit in der Evaluation. *LeGes – Gesetzgebung & Evaluation*, 23(2), 129-147.
- William, D., & Thompson, M. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 49–65). Mahwah: Lawrence Erlbaum.

Chapter 3:
**Development of the assessment for use in the evaluation of the
HAND in HAND programme⁴**

Nina Roczen, Wubamlak Endale, Mojca Rožman, Svenja Vieluf

⁴ A large part of this chapter is based on Roczen, Endale, Vieluf and Rožman (2020)

3.1. Assessment strategy for use in external evaluation of the HAND in HAND project

The HAND in HAND project was evaluated both internally and externally. This chapter presents the selection process for the instruments used in the external outcome evaluation. One focus was on the summative experiment-based evaluation of the project (see Chapter 2). For this part of the evaluation, self-report and other-report questionnaire scales, sociometric measures, and vignettes were used. The experiment based summative evaluation of the HAND in HAND programme was complemented by interviews with the evaluation of the HAND in HAND programmes from the participants' point of view. The interviews were not only used to complement our effectiveness results and to gain insights into how the programme was experienced by the participants. We also relied on them for a formative purpose, that is, we expected to learn from them how the programmes may be improved in future upscaling of the HAND in HAND programmes (for suggestions for improvement, please see Chapter 12, Chapter 12 as well as Vieluf et al., 2020).

3.2. Development process

Our development process comprised various steps, starting with a literature research, followed by a qualitative and quantitative examination of a pre-selection of questionnaire scales, and, finally, the compilation of a multifaceted measuring instrument.

3.2.1. Defining Core Concepts

Starting point for the development of the assessment, was a concrete definition of expected outcomes of the HAND in HAND programmes, developed by all project partners, and the development of a theoretical model describing the effects of the programmes on those outcomes. This provided a common basis for the development of student and school staff programmes on one hand, and the development of the instruments for the external evaluation of the programmes on the other hand, with a view to achieving the optimal alignment of both (for more information, see Kozina, Vidmar & Veldin, 2020).

The process of agreeing on the aims of the programme included intensive discussions regarding the dimension of “intercultural competence”. The overarching aim of HAND in HAND is developing “inclusive societies” – which implies overcoming “two-group-theories” (see e.g., Hinz, 2003). According to the official project title it seeks to achieve this by fostering the social, emotional

and *intercultural* (SEI) competencies of students and school staff (Kozina, Vidmar, & Veldin, this report). However, the concepts of intercultural competence and intercultural learning have been frequently criticized for essentializing culture and, thereby, cementing lines between a majority and its “migrant other” as well as for ignoring the power structures and the institutional basis of domination and discrimination (e.g., Haslam et al., 2006; Lynch, 1987; Morton et al., 2009; Prentice & Miller, 2007; Tator & Henry, 1991). This inconsistency triggered a lively and productive debate within the project team. As a result the group decided to draw on Bennett's developmental model of intercultural sensitivity (Bennett, 1986, 1993, 2004, 2014) and Blell & Doff's (2014) concept of transcultural competence instead of a more traditional concept of intercultural competence, because these concepts do not create self-other-binaries. Yet, several of the exercises used in the programmes introduced an additional perspective: they were adopted from existing diversity awareness and antiracist programmes (for a detailed description of the trainings see Jensen et al., 2018a; Jensen et al., 2018b; Marušić et al., 2020). In accordance with the content of the programmes, also most of the instruments used for evaluating the programmes measure diversity awareness instead of intercultural competence – only one questionnaire scale also addresses transcultural competence. To make clear that the original idea of fostering intercultural competence with the programmes has evolved over time and that the programme exercises and evaluation instruments mainly address diversity awareness, we will name the dimension “intercultural competencies/diversity awareness” in the following.

3.2.2. Compilation of questionnaire scales based on literature review

In step two, we researched existing open-access instruments to assess the previously defined core concepts (Denk et al., 2017). Most existing instruments targeting social and emotional competencies and intercultural competencies/diversity awareness as well as classroom climate are based on questionnaire scales that are mostly self-reports. Since several existing scales were available for each core concept in HAND in HAND, we decided to test a large number of self-report scales in a set of cognitive laboratories followed by a pilot study to underpin the selection of those for use in the evaluation, namely those with the best psychometric characteristics in the three school systems in which the HAND in HAND field trials were planned (Slovenia, Croatia, and Sweden).

3.2.3. Qualitative examination and first preselection

Prior to testing instruments in a pilot study, a set of instruments from the assessment catalogue, mostly established self-report measures and some self-developed scales, were tested in cognitive laboratories in three participating

school systems. A cognitive laboratory is a method of investigating the mental processes that take place while answering a questionnaire item (Prüfer & Rexroth, 2000). The scales to be investigated were split into three batches with each country testing one batch. For this purpose, we reached out to schools to interview a small sample of 131 students (Slovenia: 80, Sweden: 10, Croatia: 31) on the appropriateness of the selected instruments (see Table A in the Appendix for a list of tested instruments). Interviews were conducted one-on-one and lasted about two hours. During the interviews, students were asked to provide information on whether and how they understood the questions, answering options and specific terms and on why they chose a particular answering option. To ensure comparability of the process across all school systems, a protocol for contacting schools as well as for conducting and coding results of the interviews was provided. The feedback from the cognitive laboratories was used either to confirm that the instrument was appropriate for being used in the evaluation of the HAND in HAND programme, or to adapt items and answer categories and delete scales or single items (see Table A in the Appendix for an overview). For example, the scale “Group-focused enmity (generalized prejudice) measure” (Zick, Wolf, Küpper, Davidov, Schmidt & Heitmeyer, 2008) was fully deleted, because it was perceived as neither age appropriate nor culturally appropriate. Other scales were taken out because students had reported problems with understanding the items (e.g., “Social self-efficacy scale”; Muris, 2001) or because they had complained about the length of the scale (e.g., “Interpersonal Competence Questionnaire (ICQ)”; Buhrmester, Furman, Wittenberg & Reiss, 1988).

3.2.4. Selection of questionnaire scales based on pilot test

The questionnaire scales that had been pre-selected with the help of the cognitive laboratories were tested again in a quantitative pilot study and, based on the results of both pre-tests, an evaluation instrument was compiled. Section 3 of this chapter provides an overview of all types of instruments used in the evaluation. Section 4 reports in detail on the procedure and results of the pilot and presents the final compilation of evaluation instruments.

3.2.5. Selection and development of alternative measures

In parallel to researching and testing questionnaire scales, we selected and developed several other measurement types to realize a broad and multifaceted assessment strategy. These instruments encompassed other-report questionnaire scales, sociometry, interviews and vignettes.

3.3. Measures for the Evaluation

In the following, we first present the instruments we compiled to measure and understand the effects of the HAND in HAND programme on Social and emotional competencies and intercultural competencies/ diversity awareness and the classroom climate in the framework of the summative outcome evaluation. We conclude by presenting the interviews, whose purpose is twofold: On one hand, they supplement the summative outcome evaluation with the participants' perspective while, on the other hand, they provide information for use in a further development of the programmes (formative purpose of the evaluation).

3.3.1. *Measuring change in Social and emotional competencies and intercultural competencies/ diversity awareness*

Self-report questionnaire scales. With self-report scales the respondents assess themselves regarding a selected characteristic, e.g., the extent of their own aggressiveness or the ability to take another's perspectives. Even though self-reports have some deficiencies such as response biases (see e.g. Bogner & Landrock, 2015; He & Van de Vijver, 2012), they still bring several advantages such as their time-efficient and uncomplicated implementation, objectivity and comparability.

Self-report scales targeting self-awareness, self-management, social awareness, relationship skills, and intercultural competencies/diversity awareness were selected for the pilot study. The scales are shown in Table B and C in the Appendix to this chapter (see also Roczen, Endale, Vieluf, & Rožman, 2019).

Other-reports. One way to overcome some of the disadvantages of assessing competencies in the form of self-reports, such as conscious and unconscious answer tendencies, is to use "other-reports". This means that certain characteristics or competencies are not or not solely assessed by the persons concerned themselves, but the respective characteristics are (also) assessed by other persons. In the student questionnaire, we used one measure, namely the Multisource Assessment of Children's Social Competence (MASCS) (scale "Cooperation"; Junttila, Voeten, Kaukiainen, & Vauras, 2006) to compare different perspectives on students' cooperative behaviour. For each student, three randomly assigned classmates assessed that student's social behaviours, e.g., the extent to which that student offers help to others, or whether the student invites other students to participate in activities.

Vignettes. In addition to questionnaire scales, we also included a vignette (often also referred to as situational judgement tests). It starts with a brief description of a scenario, followed by questions asking the participants to assess different aspects of that scenario (Whetzel & McDaniel, 2009). The vignette is based on a situational judgement test developed by Schwarzenhal (2019) and was used in both the student and school staff questionnaire. It describes a difficult situation in the school environment that concerns a newly immigrated student and is followed by questions about the participants' interpretation of the situation and their assessment of possible behavioural options to solve the situation. Based on participants responses to the questions concerning the vignettes a coding scheme was developed that aims at classifying the way participants referred to social difference categories in their interpretation of the situation as well as the quality (students) or inclusiveness (teachers) of solutions participants had developed. The whole material was coded by at least two out of three coders who had been intensively trained before the actual coding took place. Difficult cases were discussed among coders and examples typical for a code as well as equivocal cases were noted and listed together with detailed and extensive coding rules. Interrater-reliability was computed between all pairs of coders for the inclusion-related vignette (the social perspective taking acts were coded by only one coder due to a time pressure). We chose a minimum level of interrater-reliability of 80% agreement (see Rädiker & Kuckartz, 2019). This was achieved for all codes reported in Chapter 5.

3.3.2. *Measuring change in the classroom climate*

Questionnaire scales. As for Social and emotional competencies and intercultural competencies/ diversity awareness, we also employed questionnaire scales to assess the classroom climate. Here, the participants did not assess their own competencies, but aspects of the classroom climate like the orderliness of the classroom or the relationships with their teachers.

Sociometry. Sociometry is a qualitative research technique which explores relationships among members of a group (Moreno, 1934; Wasserman and Faust, 1994). These relationships can thus be visualised in a sociogram where individuals are represented as points and the relationships between them as lines. For creating the sociograms, we adapted the approaches of Dollase (1976) and Schwab (2016) and asked students with which other students from their class they had most often spent their breaks during school over the previous 4 months and whether there were any students in their class with whom they did not spend any of their breaks during that time. Indicators for the quality of the classroom climate we derive from the answers to these questions were based on

suggestions made in Hennig, Brandes, Pfeffer, & Mergel (2012) and Jansen (2006). We computed (a) the density of the social network in the classroom (number of reported relations divided by the number of possible relations in a class); (b) the percentage of isolated students (students who did not report to spend their breaks commonly with any other students from the class) and (c) the percentage of unpopular students (students that were named by no other students or by only one other student in response to the question with whom they commonly spent their breaks).

3.3.3. Semi-structured focus-group interviews

With the focus group interview method, groups of individuals are guided by questions such that they can interact with each other and give responses that are related to the contributions of other participants (e.g. Vaughn, Schumm, & Sinagub, 1996). Interviews are a particular important component of an evaluation as they consider the perspectives of participants (for more details, see Vieluf et al., 2020). Three different group-interviews (students, teachers and school leaders together with other school staff) took place in each participating school in all three school systems. The interviews took place between 3 and 6 months after the trainings had been finished. The interviews were done in the schools. They were done by the HAND in HAND partners who were responsible for implementing the programmes in Slovenia, Sweden and Croatia, respectively. The persons who conducted workshops with students interviewed the teachers and leadership, and the ones who conducted the workshops with school staff did the interviews with the students. The three control schools were interviewed by the colleagues who did not conduct any workshops. The timeframe was 45 minutes per group. There were always two persons present. Questions from the interview guide encompass why the school had taken part in the HAND in HAND programme (only school staff), how they liked the programme and particular exercises, what they had learned from them, whether they had any suggestions to help improve the programmes, whether they were still practising some of the exercises themselves and whether they had noticed positive (perhaps also negative) changes in the classroom climate or in their teachers (only students).

Responses to the interviews were analysed by means of qualitative content analysis (e.g., Schreier, 2012). The first step was gaining an overview over the material and marking relevant sections. The second step was developing a coding system. For most of the questions we used inductive coding. Only for the analysis of responses to two questions – what participants learned through the HAND in HAND training and what they would highlight as perceived outcomes – the definition of categories was theory-driven (deductive). More specifically, the

codes were derived from the definition of core constructs described in Chapter 1. All codes (inductive as well as deductive) were, in the third step, applied to the whole material. Finally, it was counted how often each code was ascribed to an interview-answer and quotes from the interviews were selected to illustrate some of the codes.

The interview data had the purpose to move beyond detecting possible positive or negative effects of the programme. The interviews allowed us to understand how the participants experienced the programme. They also served a formative purpose and gave us suggestions for how to improve the programmes from the participants' perspectives.

3.4. Selection of Questionnaire Scales – Pilot Study

While tests and qualitative instruments addressing Social and emotional competencies and intercultural competencies/ diversity awareness as well as classroom climate are quite hard to find, many questionnaire scales addressing the HAND in HAND core constructs are available (Denk et al., 2017). To help selecting from among these scales we used the following procedure: First, we made an extensive and systematic review of the literature describing self-report scales that assess the core concepts of the HAND in HAND programme (Denk et al., 2017). From this collection, we chose several alternative instruments assessing each respective core construct. To help select between those scales measuring the same construct, the scales were presented to the students and the teachers in the HAND in HAND pilot study. The methods and results of that study are described below.

3.4.1. Methods

Participants. For the pilot study, we collected data from convenience samples at schools in Sweden, Croatia and Slovenia. The target group was 13- to 14-year-olds (grade 8 students) and their teachers⁵. A summary of the students' and teachers' demographic characteristics is given in Table 3.1.

The average age of the students was 13.2 years in Slovenia, 14.0 years in Croatia and 14.7 years in Sweden. The share of girls in percent was 29.4 % in Croatia, 51.9 % in Slovenia and 53.5 % in Sweden. In the Slovenian sample, 1.6 % of the students were born outside of Slovenia, 3.3 % usually speak a language other than

⁵ In the Field Trials and in the Field Trial data collections, not only teachers, but also school principals, school social workers and counsellors were addressed. In the pilot study, questionnaires were only handed to teachers.

Slovenian at home and another 6.6 % usually speak Slovenian and (an)other language(s) at home. In Sweden, 3.2 % of the students were born outside of the country, 1.1 % usually speak a language other than Swedish at home and another 6.5 % usually speak Swedish and (an)other language(s) at home. In the Croatian Sample, 1 % of the students were born outside of Croatia, 1 % usually speak a language other than Croatian at home and another 8.8 % usually speak Croatian and (an)other language(s) at home.

The teachers' average age in Croatia was 42.6 years, 43.2 in Sweden and 44.1 in Slovenia. In Sweden, 80.4 % of the teachers were female, in Croatia 88.6 % and in Slovenia 89.6 %. In Slovenia, 3.1 % of the teachers were born outside of the country, in Sweden, the percentage was 5.3 % and in Croatia 22.8 %.

Table 3.1.

Pilot study sample sizes and demographic characteristics of the students and teachers

		Students				Teachers			
		<i>Overall</i>	<i>Slovenia</i>	<i>Sweden</i>	<i>Croatia</i>	<i>Overall</i>	<i>Slovenia</i>	<i>Sweden</i>	<i>Croatia</i>
Sample Size	<i>N</i>	623	234.0	181	208	284	97	93	78
Age	<i>M (SD)</i>	13.9 (0.83)	13.2 (0.43)	14.7 (0.84)	14.0 (0.37)	43.2 (9.61)	44.1 (9.55)	43.2 (9.79)	42.6 (9.41)
Gender	<i>Female (%)</i>	44.6	51.9	53.5	29.4	85.4	89.6	80.4	88.6
	<i>Male (%)</i>	55.1	48.1	45.3	70.6	14.6	10.4	19.6	11.4
	<i>Diverse (%)</i>	0.3	0.0	1.2	0.0	n.a.	n.a.	n.a.	n.a.
Country of Birth	<i>“Other“ country (%)</i>	1.9	1.6	3.2	1.0	9.2	3.1	5.3	22.8
Language at Home	<i>Language of the country (%)</i>	90.5	90.2	92.5	89.2	n.a.	n.a.	n.a.	n.a.
	<i>Other language(s) (%)</i>	1.9	3.3	1.1	1.0	n.a.	n.a.	n.a.	n.a.
	<i>Both (%)</i>	7.6	6.6	6.5	9.8	n.a.	n.a.	n.a.	n.a.

Note. *M* = Mean, *SD* = standard deviation. Information on the students' gender was only collected from those students who answered booklet B. Therefore, information on the students' gender is only based on *N* = 297 students. In the teacher questionnaire, the question on gender only included two options. Information on the country of birth and language spoken at home was only collected from those students who answered booklet A (*N* = 326).

Measures. The full pilot study instrument for students encompassed 31 scales covering students' self-reported Social and emotional competencies and intercultural competencies/ diversity awareness as well as their perception of the classroom climate. For the student data collection, we used two booklets to test a larger number of instruments and remain time efficient. Each student was presented with one booklet so that each item was only answered by about half of

the students. The pilot questionnaire for the teachers included 23 scales covering the teachers' self-reported Social and emotional competencies and intercultural competencies/ diversity awareness as well as their perception of the classroom climate (see Table C in the Appendix to this chapter). Fourteen scales were used in both the student and teacher questionnaires (see the column "Parallel scale in SSQ" in Table B in the Appendix to this chapter and "Parallel scale in TCQ" in Table C in the Appendix to this chapter).

Procedures. We performed the following analyses to ensure the aforementioned criteria were available for scale selection: We analysed (i) descriptive statistics on the item level (frequencies and missing values), (ii) descriptive statistics on the scale level (scale means and standard deviations), (iii) the dimensionality of the scales using exploratory factor analyses (EFA) and (iv) the internal consistency of scales (Cronbach's alpha). We analysed data for each country separately. All statistical analyses were carried out using IBM SPSS Statistics Version 22 for Windows.

3.4.2. Results

The results of the analysis of the distributions and percentages of missing values, internal consistency and exploratory factor analysis are shown in Table B in the Appendix to this chapter for the student data and in Table C in the Appendix to this chapter for the teacher data.

Students. Overall, the internal consistencies of the scales in the student questionnaire (see Table B in the Appendix to this chapter) are reasonable (DeVellis, 2003). For about half the scales (15), the reliability is above $\alpha = .85$ in at least one country. For two-thirds of the scales (20 out of 31 scales), the reliabilities in all three school systems are above $\alpha = .70$. For seven other scales, the reliability is at least $\alpha = .60$.

The number of missing values is acceptable in all school systems for most of the scales in the student questionnaire, i.e. < 10% in 17 out of 31 scales. For most scales, the number of missing responses is lower in Croatia and Slovenia than in Sweden. While the percentage of missing values lies between 0% and 2% for a large part of the scales in Croatia and Slovenia, a considerable range is observed in Sweden. For example, for five scales, less than 5% of responses are missing for the single items, but for 10 scales, there are up to 15%–25% missing values. These results show that many of the Swedish participants did not complete their questionnaire.

As regards the distributions, the mean values of positively worded scales are generally relatively high. The scales with the highest mean values (with respect to

the possible maximum value) are “Prosocial behaviour” ($M = 3.28 - 3.55$), “Orderliness of the classroom” ($M = 3.20 - 3.36$) and “Inclusive classroom climate” ($M = 2.98 - 3.29$). Hence, positively worded scales – and the latter scales in particular – are skewed.

For the lion’s share of the scales (21 out of 31), the factor structure is identical across the school systems (see “✓” in the “EFA” column in Table B in the Appendix to this chapter). For the remaining scales, the number of extracted factors differs between school systems. In most cases, a scale is one-dimensional as theoretically anticipated in some school systems, whereas it is two-dimensional in others. Items usually group into two factors where one is characterized by the positively worded items and the other by the negatively worded ones.

Teachers. The reliabilities of the teacher scales are good or very good (see Table C in the Appendix to this chapter). For almost all scales (21 out of 23 scales), the reliabilities in all three school systems are above $\alpha = .70$. For 16 of the scales, the reliability is $\alpha = .85$ or above in at least one country.

As in the student sample, missing values in Croatia and Slovenia are very low (often even 0%). In Sweden, the number of missing teacher responses is much higher and also considerably higher than in the Swedish student data set. Again, a wide range of missing values can be observed. For the scale “Observe” of the “Kentucky Inventory of Mindfulness Skills” there are 18.9% of missing values, while one item of the scale “Professional Beliefs about Diversity Scale” is even missing for all participants. Here, too, these results show that many participants did not complete the questionnaire.

As with the student scales, the mean values of positively worded scales are relatively high. The scales with the highest mean values (relative to the possible maximum value) are “Teacher Self-Efficacy” ($M = 3.13 - 4.13$), “Empathic concern” ($M = 3.84 - 4.07$) and “Reflexivity” ($M = 3.22 - 3.36$).

For about half the scales (13 out of 23), the dimensionality is consistent across the school systems (see the “EFA” column in Table C in the Appendix to this chapter).

3.4.3. Selection of scales for Field Trial

While selecting one out of two or more scales intended to measure a similar construct, we applied the following criteria: (1) the accuracy with which one scale measures a construct (i.e. internal consistency – Cronbach’s α ; we regarded values above $\alpha = .70$ as acceptable); (2) the correspondence of the number of extracted factors with the theoretically expected dimensionality in all school

systems – as this is a necessary precondition that has to be given if data analysis across school systems or country comparisons are intended. We also checked (3) the distribution of the participants' responses – we primarily looked at those to identify ceiling effects as it is difficult to detect possible programme effects with instruments that are already strongly skewed in the direction of the expected effects. For these first three criteria, we used results from the pilot study described above. We also considered (4) the efficiency of a scale in terms of the expected response time. As an indicator for this efficiency, we relied on the number of items per scale but also an estimation of the response time that we determined in individual trial runs outside of the pilot survey. In addition to applying these criteria, we ensured that the entire range of HAND in HAND core concepts (see Chapter 1) was covered by the scales selected for the summative evaluation. To sum up, we chose the scale that was ideally more reliable, had a less skewed distribution, had the same structure in the three school systems, and was shorter than the other scales. In many cases, the competing scales performed well in different analyses, making it sometimes difficult to choose the more suitable one. In these cases, we prioritised the selection criteria according to the above numbering (criterion no. 1 was the most important and criterion no. 4 the least important to be considered). The requirement to consider all core constructs led to the inclusion of a few scales that did not perform optimally. For example, the scale "Self-Awareness" shows unsatisfactory reliability in the Slovenian student sample and also the dimensionality was not consistent across the school systems (see Table B in the Appendix to this chapter). However, since practising self-awareness is a fundamental core concept of the HAND in HAND programme (see Kozina et al., 2020; Jugović, Puzić and Mornar, 2020; Jensen and Gøtzsche, 2020), we nevertheless decided to keep the scale. This and similar scales are examined particularly critically in the analysis of the Field Trial data. In a few cases, we decided to shorten the scales (see Table B in the Appendix to this chapter, the "# Items Field Trial" column). For instance, for the "Inclusive Classroom Climate" scale we kept only the negatively worded items to ensure a consistent structure in all school systems.

3.5. Conclusion - Final Evaluation Instruments for the HAND in HAND Field Trials

All instruments included in the final Field Trial evaluation questionnaire are listed in the "Measures" column (printed in black) in Table B and Table C in the Appendix to this chapter.

In the process of developing the assessment for external evaluation of the HAND in HAND programme, our assessment strategy had the following characteristics:

(1) optimal alignment between the HAND in HAND programmes and the evaluation instruments by reference to common core concepts; (2) a multi-method approach to take account of both processes and outcomes and to capture different levels on which effects may occur; and (3) the pre-testing of a large part of the instruments in order to have measures available that are equally well suited for use in all participating school systems. This should establish optimal conditions for measuring and explaining the effectiveness of the HAND in HAND programme and for providing data that can be used to optimise it.

3.6. References

- Baer, R. A., Smith G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky inventory of mindfulness skills. *Assessment, 11*(3), 191–206.
- Belmont, M., Skinner, E., Wellborn, J., Connell, J., & Pierson, L. (1992). *Teacher as social context (TASC): Two measures of teacher provision of involvement, structure, and autonomy support*. Unpublished manuscript, Graduate School of Education and Human Development, Department of Psychology, University of Rochester.
- Bogner, K. & Landrock, U. (2015). *Antworttendenzen in standardisierten Umfragen*. Mannheim, GESIS Leibniz Institut für Sozialwissenschaften (GESIS Survey Guidelines). Retrieved from: <https://www.gesis.org/gesis-survey-guidelines/instruments/fragebogenkonstruktion/antworttendenzen/>
- Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology 3*, 77-101.
- Buhrmester, D., Furman, W., Wittenberg, M. T., & Reis, H. T. (1988). Five domains of interpersonal competence in peer relationships. *Journal of Personality and Social psychology, 55*(6), 991.
- Cornell, D. (2016). The Authoritative School Climate Survey and the School Climate Bullying Survey: Research summary. 10.13140/RG.2.1.4948.2325.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalogue of Selected Documents in Psychology, 10*, 85.
- Denk, A., Müller, F. Lubaway, E., Sälzer, C., Kozina, A., Vidmar, M., Rasmus-

- son, M., Marusic, I., Jensen, H., Nielsen, B. L., Roczen, N., Ojsteršek, A. & Jurko, S. (2017). Catalogue for the Assessment of Social, Emotional, and Intercultural Competencies.
- Diazgranados, S., Selman, R. L., & Dionne, M. (2016). Acts of social perspective taking: A functional construct and the validation of a performance measure for early adolescents. *Social Development*, 25(3), 572–601.
- Diemer, M. A., Rapa, L. J., Park, C. J., & Perry, J. C. (2017). Development and validation of the Critical Consciousness Scale. *Youth & Society*, 49(4), 461–483.
- Dollase, R. (1976). *Soziometrische Techniken: Techniken der Erfassung und Analyse zwischenmenschlicher Beziehungen in Gruppen*. Weinheim: Beltz.
- Elias, S. M., & Loomis, R. J. (2000). Using an academic self-efficacy scale to address university major persistence. *Journal of College Student Development*, 41, 450–454.
- Geldhof, G. J., Bowers, E. P., Boyd, M. J., Mueller, M. K., Napolitano, C. M., Schmid, K. L., & Lerner, R. M. (2014). The creation and validation of short and very short measures of PYD. *Journal of Research on Adolescence*, 24(1), 163–176.
- Gilman, R., Laughlin, J. E., & Huebner, E. S. (1999). Validation of the Self-Description Questionnaire-II with an American sample. *School Psychology International*, 20(3), 300–307.
- Goodenow, C. (1993). The psychological sense of schoolmembership among adolescents: scale development and educational correlates. *Psychology in the Schools*, 30(1), 79-90.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586.
- Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry*, 7(3), 125–130.
- Fischer, N., Decristan, J., Theis, D., Sauerwein, M. & Wolgast, A. (2017). *Skalendokumentation (online): Studie zur Entwicklung von Ganztags- schulen - Teilstudie StEG-S*, In: Datenbank zur Qualität von Schule (DaQS), Frankfurt am Main: Deutsches Institut für

- Internationale Pädagogische Forschung (DIPF). Retrieved from: <https://daqs.fachportal-paedagogik.de/search/show/survey/136>
- He, J., & Van de Vijver, F. J. R. (2012). Bias and equivalence in cross-cultural research. *Online readings in Psychology and Culture*, 2(2), 8. Retrieved from: <https://doi.org/10.9707/2307-0919.1111>
- Hennig, M., Brandes, U., Pfeffer, J., & Mergel, I. (2012). *Studying Social Networks : A Guide to Empirical Research*. Frankfurt [u.a.]: Campus Verlag.
- Jansen, D. (2006). *Einführung in die Netzwerkanalyse: Grundlagen, Methoden, Forschungsbeispiele* (3. Auflage). Bielefeld: VS Verlag.
- Jensen, H., & Gøtzsche, K. (2020). Development of the social, emotional and intercultural learning programme for school staff. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (83-105). Hamburg: Verlag Dr. Kovač.
- Jugović, I. Puzić, S., & Mornar, M. (2020). Development of the social, emotional and intercultural learning programme for students. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (59-82). Hamburg: Verlag Dr. Kovač.
- Junttila, N., Voeten, M., Kaukiainen, A., & Vauras, M. (2006). Multisource assessment of children's social competence. *Educational and Psychological Measurement*, 66(5), 874–895.
- Kaukiainen, A., Björkqvist, K., Österman, K., Lagerspetz, K., & Forsblom, S. (1995). *Peer-estimated social intelligence (PESI)*. Turku, Finland: Department of Psychology, University of Turku.
- Kaukiainen, A., Björkqvist, K., Lagerspetz, K., Österman, K., Salmivalli, C., Rothberg, S., & Ahlbom, A. (1999). The relationships between social intelligence, empathy, and three types of aggression. *Aggressive Behaviour*, 25(2), 81–89.
- Kozina, A. (2013). The LA aggression scale for elementary school and upper secondary school students: examination of psychometric properties of a new multidimensional measure of self-reported aggression. *Psihologija*, 46(3), 245–259.

- Kozina, A. (2012). The LAOM multidimensional anxiety scale for measuring anxiety in children and adolescents: Addressing the psychometric properties of the scale. *Journal of Psychoeducational Assessment*, 30(3), 264-273.
- Kozina, A., Vidmar, M., & Veldin, M. (2020). Social, emotional and intercultural/transcultural learning in a European perspective: Core concepts of the HAND in HAND project. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (17-37). Hamburg: Verlag Dr. Kovač.
- Marsh, H. W. (1990). *Self-Description Questionnaire-III: SDQ III Manual*. University of Western Sydney, Macarthur.
- Moreno, J. L. (1934) *Who shall survive? A new approach to the problem of human interrelations*. New York: Beacon House.
- Muris, P. (2001). A brief questionnaire for measuring self-efficacy in youths. *Journal of Psychopathology and Behavioural Assessment*, 23(3), 145– 149.
- OECD (2018). *PISA 2018 Global Competence Questionnaire*. Retrieved from: <http://www.oecd.org/pisa/PISA-2018-Global-Competence-Questionnaire.pdf>
- OECD (2018). *PISA 2015 technical report*. Paris: OECD.
- Olweus, D. (1996). *The revised Olweus bully/victim questionnaire*. University of Bergen, Research Center for Health Promotion.
- Olweus Sample School Report (2007). Olweus Bullying Questionnaire: Standard School Report. Retrieved from: https://www.pdstats.com/PublicFiles/Olweus_Sample_Standard_School_Report.pdf.
- Prüfer, P. & Rexroth, M. (2000). Zwei-Phasen-Pretesting. *ZUMA-Arbeitsbericht 2000/08*. Mannheim: ZUMA.
- Rädiker, S. & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA. Text, Audio und Video*. Wiesbaden: Springer VS
- Roczen, N., Endale, W., Vieluf, S., & Rožman, M. (2020). Development of the assessment for use in evaluation of the HAND in HAND programme. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school*

environments across Europe (131-156). Hamburg: Verlag Dr. Kovač.

- Roczen, N.*, Rožman, M.*, Vieluf, S.*, Lalla, J., Dahlström, H., Lund Nielsen, B., Kozina, A., Müller, F., Puzić, S. (2019). *Evaluation of the HAND in HAND Programme - Documentation of the Questionnaire Scales*. Available at: www.handinhand.si
- Sangalang, C. C., Chen, A. C., Kulis, S. S., & Yabiku, S. T. (2015). Development and validation of a racial discrimination measure for Cambodian American adolescents. *Asian American Journal of Psychology*, 6(1), 56–65.
- Schulz, W., Ainley, J., & Fraillon, J. (Eds.) (2011). *ICCS 2009 technical report*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: Sage.
- Schwab, S. (2016). Erfassung von sozialer Partizipation–Übereinstimmung zwischen Selbst- und Fremdsicht von Schülern mit und ohne sonderpädagogischen Förderbedarf. *Zeitschrift für Pädagogische Psychologie*, 30(4), 227–236.
- Schwarzenthal, M. (2019). *Intercultural competence among adolescents attending culturally diverse schools in Germany (Unpublished Dissertation)*. University of Potsdam, Potsdam, Germany.
- Stewart-Brown, S. L., Platt, S., Tennant, A., Maheswaran, H., Parkinson, J., Weich, S., Tennant, R., Taggart, F., & Clarke, A. (2011). The Warwick-Edinburgh mental well-being scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects. *Journal of Epidemiology and Community Health*, 65(2), 38–39.
- Stöber, J. (2002). Skalendokumentation „Persönliche Ziele von SchülerInnen“. In Dalbert, C. (Ed.), *Hallesche Berichte zur Pädagogischen Psychologie*. Halle (Saale): Martin-Luther-Universität Halle-Wittenberg, Institut für Pädagogik. Retrieved from: https://daqs.fachportal-paedagogik.de/search/show/instrument/1256_32
- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2), 271-324.

- Valois, R. F., & Zullig, K. J. (2013). Psychometrics of a brief measure of emotional self-efficacy among adolescents from the United States. *Journal of school health, 83*(10), 704-711.
- Vieluf, S., Denk, A., Rožman, M., & Roczen, N. (2020). How do the participants evaluate the HAND in HAND programme? Results of semi-structured focus group interviews In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (195-218). Hamburg: Verlag Dr. Kovač.
- Vaughn, S., Schumm, J. S., & Sinagub, J. M. (1996). *Focus group interviews in education and psychology*. Thousand Oaks, CA, US: Sage Publications.
- Vidmar, M. & Kerman, K. (2016). The development of Teacher's relational competence scale: Structural validity and reliability. *Šolsko Polje, 27*(1/2), 41– 62.
- Wasserman, S. & Faust, K. (1994) *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge University Press.
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*(3), 188–202.
- Williams, D.R., Yu, Y., Jackson, J.S., & Anderson, N.B. (1997). Racial Differences in Physical and Mental Health: Socioeconomic Status, Stress, and Discrimination. *Journal of Health Psychology, 2*(3), 335-351.
- Zick, A., Wolf, C., Küpper, B., Davidov, E., Schmidt, P., & Heitmeyer, W. (2008). The syndrome of group-focused enmity: The interrelation of prejudices tested with multiple cross-sectional and panel data. *Journal of Social Issues, 64*(2), 363-383.

Appendix

Table A. Instruments tested in cognitive laboratories.

INSTRUMENT	AUTHORS	INSTRUMENT KEPT FOR PILOT STUDY			INSTRUMENT DELETED	
		<i>kept completely</i>	<i>kept partly</i>	<i>adaptations to answering categories</i>	<i>quality reasons</i>	<i>other reasons</i>
Brief Self-Control Scale	Tangney, Baumeister & Boone (2004)	✓		✓		
Attitudes towards immigrants measure	Schulz, Ainley, & Fraillon (2011)	✓				
Critical Consciousness Scale	Diemer, Rapa, Park & Perry (2017)		✓	✓		
Teacher Fairness	OECD (2018)					✓
Strength and Difficulties Questionnaire	Goodman (1997); Goodman et al. (1998)		✓	✓		
LA aggression scale for children and adolescents	Kozina (2013)	✓		✓		
Interpersonal Competence Questionnaire	Buhrmester et al. (1988)				✓	
Academic Self-Efficacy Scale	Elias & Loomis (2000)					✓
Group-focused enmity (generalized prejudice) measure	Zick et al. (2008)				✓	
Adolescent Discrimination Distress Index	Sangalang, Chen, Kulis & Yabiku (2015)		✓			
Everyday Discrimination Scale	Williams, Yu, Jackson & Anderson (1997)				✓	
Social self-efficacy scale	Muris (2001)				✓	
Interpersonal reactivity index	Davis (1980)		✓			
Olweus Bully/Victim Questionnaire	Olweus, D (1996)		✓		✓	
Short Warwick-Edinburgh Mental Well-Being Scale	Stewart-Brown et al. (2011)	✓				
Positive Youth Development Questionnaire	Geldhof et al. (2014); Lerner et al. (2005)		✓	✓		
Sense of school membership	Goodenow (1993)				✓	
Teacher as Social Context	Belmont, Skinner, Wellborn, & Connell (1992)				✓	
Emotional self-efficacy	Valois & Zullig (2013)				✓	
Scale "Bullying by Teachers" from the "Authoritative School Climate Survey"	Cornell (2016)				✓	
Perceived quality of student-teacher relations (positively worded)/"Teacher recognition"	Fischer, Decristan, Theis, Sauerwein & Wolgast (2017)		✓	✓		
Inclusive Classroom Climate	OECD (2018)		✓			
LAOM anxiety scale for children and adolescents	Kozina, A (2012)				✓	
General self-concept scale	Marsh 1990	✓				

Note. Some instruments were deleted based on the cognitive laboratories results (see column "quality reasons"), e.g. because they had caused comprehension problems or because they were classified as culturally inappropriate or offensive. Other instruments were deleted after cognitive laboratories without having received negative evaluations (see column "other reasons"). The reasons for a deletion were, for example, a conceptual overlap with other instruments or a low relevance with regard to HAND in HAND core concepts.

Table B. Overview of instruments for the Student Questionnaire (STQ)

Measures	# Items	# Items Field Trial	Pilot	Range	Parallel scale in SSQ	Time	Cronbach's α	EFA	Missings (%)	Distribution M (SD)
Self-Awareness										
Short Warwick-Edinburgh Mental Well-Being Scale (Stewart-Brown et al., 2011)	14	-	✓	1-4		01:29	.79-.86	✓	0.9 (SVN) - 3.8 (HRV)	3.06 (0.39) - 3.18 (0.44)
Positive Youth Development Questionnaire (Geldhof et al., 2014; Lerner et al., 2005) – Scale Positive identity	6	6	✓	1-4		00:50	.72-.87	✓	1.2 (HRV) - 6.3 (SVN)	2.81 (0.53) - 3.05 (0.59)
Self-Description Questionnaire (SDQ-II; Gilman, Laughlin & Huebner, 1999; Marsh, 1990) - General Self-concept	10	-	✓	1-4		01:16	.74-.93	✓	1.1 (SWE) - 3.9 (SVN)	3.02 (0.39) - 3.18 (0.49)
<i>Kentucky Inventory of Mindfulness Skills (Baer, Smith & Allen, 2004):</i>										
Scale 'Describe'	8	7	✓	1-5	x	01:03	.52-.84	-	1.2 (HRV) -12.6 (SWE)	2.97 (0.55) - 3.44 (0.74)
Scale 'Accept without Judgement'	9	9	✓	1-5	x	01:14	.83-.87	✓	0.9% (SVN) – 12.8% (SWE)	2.56 (0.81) – 2.96 (0.75)
Scale 'Act with awareness'	10	10	✓	1-5	x	01:27	.74-.79	-	0.9 (SVN)-12.8 (SWE)	2.96 (0.66) - 3.04 (0.60)
Scale 'Observe'	12	7	✓	1-5	x	01:15	.71-.91	✓	1.2 (HRV) - 11.6 (SWE)	2.80 (0.83) - 3.31 (0.71)
Self-Management										
Self-Efficacy Questionnaire for Children (SEQ-C; Muris, 2001) - Emotional Self-Efficacy (ESE)	8	-	✓	1-5	x	01:39	.72-.87	✓	0.9 (SVN) - 3.4 (SWE)	3.04 (0.96) - 3.39 (0.65)
<i>Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; Goodman, Meltzer & Bailey, 1998)</i>										
Hyperactivity Scale	6	-	✓	1-4		00:44	.63-.82	-	0.9 (SVN) - 4.7 (SWE)	2.42 (0.65) - 2.22 (0.58)
Emotional Problems	7	7	✓	1-4	x	00:47	.79-.84	✓	0.9 (SVN) - 3.5 (SWE)	2.24 (0.62) - 2.38 (0.68)
Brief Self-control Scale (BSCS; Tangney, Baumeister & Boone, 2004) - Scale self-management	13	-	✓	1-4		01:30	.72-.80	✓	0.9 (SVN) - 5.8 (SWE)	2.49 (0.49) - 2.59 (0.38)
LA aggression Scale (LAS; Kozina, 2013)	18	18	✓	1-4		00:42	.84-.89	✓	1.2 (HRV) - 7.4 (SWE)	1.99 (0.42) - 2.12 (0.51)
Social-Awareness										
<i>Interpersonal Reactivity Index (IRI; Davis, 1980) - Multidimensional assessment of Empathy:</i>										
Scale 'Fantasy'	7	-	✓	1-4	x	01:15	.63-.74	✓	0.9 (SVN) – 7.0 (SWE)	2.47 (0.63) - 2.57 (0.63)
Scale 'Empathic concern'	7	7	✓	1-4	x	01:15	.61-.76	-	0.9 (SVN) – 7.0 (SWE)	2.66 (0.52) - 2.97 (0.49)
Scale 'Perspective taking'	7	7	✓	1-4	x	00:50	.66-.79	✓	0.9 (SVN) – 7.0 (SWE)	2.58 (0.64) - 2.62 (0.53)
Scale 'Personal distress'	7	-	✓	1-4	x	00:50	.68-.78	✓	0.9 (SVN) - 8.1 (SWE)	2.14 (0.61) - 2.36 (0.56)
Relationship skills										
Positive Youth Development Questionnaire (PYDQ; Geldhof et al., 2014) – Scale Caring	9	9	✓	1-4		00:30	.86-.91	✓	1.2 (HRV) - 5.5 (SVN)	2.97 (0.57) - 3.22 (0.63)
<i>Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; Goodman et al., 1998):</i>										
Scale 'Prosocial behaviour'	6	-	✓	1-4		01:20	.78-.83	✓	0 (HRV) - 5.3 (SWE)	3.28 (0.58) - 3.55 (0.49)
Scale 'Peer Relationship Problems'	6	-	✓	1-4		00:36	.51-.59	-	1.2 (HRV) - 9.5 (SWE)	1.87 (0.39) - 1.98 (0.47)
Scale 'Peer Problems'	5	-	✓	1-4		00:21	.56-.64	✓	2.4 (HRV/SVN) - 8.4 (SWE)	1.78 (0.51) - 1.94 (0.51)
Self-Efficacy Questionnaire for Children (SEQ-C; Muris, 2001) - Social Self-Efficacy Scale (SSE)	8	-	✓	1-5		01:10	.65-.72	✓	0.9 (SVN) - 9.3 (SWE)	3.58 (0.66) - 3.69 (0.54)
Peer-Estimated Social Intelligence and Empathy (PESI/PEE; Kaukiainen, Björkqvist, Österman, Lagerspetz & Forsblom, 1995; Kaukiainen et al., 1999)	19	-	✓	1-4		01:20	.84-.94	✓	1.2 (HRV) - 21.1 (SWE)	2.78 (0.64) - 2.99 (0.41)
Other report: Multisource Assessment of Children's Social Competence (MASCS; Junttila, Voeten, Kaukiainen & Vauras, 2006) – Scale 'Cooperation'		5*3		1-4		10:00				
Classroom Climate										
Orderliness of the classroom (OOC-S; OECD, 2005)	5	5	✓	1-4		00:50	.77-.88	✓	0.9 (SVN) - 15.1 (SWE)	3.20 (0.70) - 3.36 (0.70)
Teacher as Social Context (TASC, 1992) Belmont, M., Skinner, E., Wellborn, J., Connell, J., & Pierson, L. (1992)	9	-	✓	1-4		01:35	.90-.91	✓	1.2 (HRV) - 14.7 (SWE)	2.35 (0.69) - 2.58 (0.65)
Perceived quality of student-teacher relations (positively worded; Fischer, Decristan, Theis, Sauerwein & Wolgast, 2017)	8	8	✓	1-4		00:53	.87-.94	✓	1.3% (HRV) – 16.3% (SWE)	2.65 (0.63) – 3.22 (0.86)
Perceived quality of student-teacher relations (negatively worded; OECD, 2018)	7	7	✓	1-4		00:44	.63-.83	✓	0.0% (HRV) – 14% (SWE)	1.82 (0.83) – 1.95 (0.72)
Inclusive Classroom Climate (ICC; OECD, 2018)	7	4	✓	1-4		00:54	.75-84	-	1.2 (HRV) - 17.9 (SWE)	2.98 (0.52) - 3.29 (0.58)
Social Climate in the Classroom (SCC; Stöber, 2002)	10	-	✓	1-4		01:00	.68-79	-	1.3 (HRV) - 15.1 (SWE)	2.49 (0.48) - 2.57 (0.39)
Olweus Bully/Victim Questionnaire (OBVQ; Olweus, 1996; Olweus Sample School Report, 2007)	22	-	✓	1-4		01:03	.85-.92	✓	1.2 (HRV) - 15.8 (SWE)	1.20 (0.36) - 1.34 (0.42)
Adolescent Discrimination Distress Index (ADDI; Sangalang, Chen, Kulis & Yabiku, 2015)	7	7	✓	1-4		00:33	.77-.90	-	1.3 (HRV) - 17.4 (SWE)	1.61 (0.73) - 1.84 (0.83)
Teachers' Relational Competence Scale (TRCS; Vidmar & Kerman, 2016)	11	9	✓	1-4	x	01:13	.80-91	✓	1.2 (HRV) - 21.1 (SWE)	2.78 (0.63) - 2.99 (0.44)
Sociometric Measure (adapted from Dollase, 1976, and Schwab, 2016)	-	2		n.a.		05:00				
Inter(trans)cultural competencies										
Attitudes Towards Immigrants (Schulz, Ainley, & Fraillon, 2011)	6	6	✓	1-4		01:12	.79-.87	✓	1.3 (HRV) - 18.6 (SWE)	2.81 (0.69) - 2.99 (0.79)
Critical Consciousness Scale (CCS; Diemer, Rapa, Park & Perry, 2017)	10	-	✓	1-4	x	01:07	.46-.58	✓	1.3 (HRV) - 18.6 (SWE)	2.35 (0.56) - 2.59 (0.58)
Vignette: Intercultural Awareness (Schwarzenthal et al., 2017)	-	3		n.a.	x	10:00				
Measures targeting several areas										
Focus group interviews	n.a.	n.a.		n.a.	x					

Note. Instruments printed in black font in the "measures" column are part of the final Field Trial questionnaires. The "pilot" column indicates which of those instruments were tested in the Pilot Study. Instruments or single scales appearing in grey font were excluded after the Pilot Study. In the "range" column, the possible response range for each scale is displayed so that the mean values in the "distribution M (SD)" column can be interpreted in relation to it. The values in the "missings (%)" column refer to individual items within a scale: For each scale, a percentage is given for the item with the lowest number of missings and for the item with the highest number of missings (each in a country comparison). The "EFA" column shows whether the factor solution, i.e. the structure across the school systems was comparable (= "✓") or inconsistent (= "-").

Table C. Overview of instruments for the School Staff Questionnaire (SSQ).

Measures	# Items	# Items Field Trial	Pilot	Range	Parallel scale in STQ	Time	Cronbach's α	EFA	Missings (%)	Distribution $M (SD)$
Self-Awareness										
<i>Kentucky Inventory of Mindfulness Skills (Baer et al., 2004):</i>										
Scale Describe	8	7	✓	1 - 5	x	01:03	.73-.93	✓	0.0 (HRV) - 22.1 (SWE)	3.51 (0.55) - 3.85 (0.60)
Scale Accept without Judgement	9	9	✓	1 - 5	x	01:14	.86-.90	✓	1.3 (SVN) - 28.4 (SWE)	3.28 (0.78) - 3.60 (0.71)
Scale Act with awareness	10	10	✓	1 - 5	x	01:27	.75-.83	-	0.0 (HRV) - 28.4 (SWE)	3.08 (0.50) - 3.57 (0.49)
Scale Observe	12	7	✓	1 - 5	x	01:15	.87-.90	✓	0.0 (HRV) - 18.9 (SWE)	3.49 (0.69) - 3.60 (0.62)
Teacher Self-Efficacy (TSE; OECD, 2013a)	12	-	✓	1 - 5		02:13	.70-.87	✓	0.0 (HRV) - 69.5 (SWE)	3.13 (0.30) - 4.13 (0.42)
Self-Management										
Self-Efficacy Questionnaire for Children (SEQ-C; Muris, 2001) – Scale Emotional Self-Efficacy (ESE)	8	-	✓	1 - 5	x	01:39	.74-.86	✓	0.0 (HRV) - 32.6 (SWE)	3.28 (0.63) - 3.45 (0.52)
<i>Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; Goodman, Meltzer & Bailey, 1998):</i>										
Scale 'Hyperactivity'	6	-	✓	1 - 5	x	00:44				
Scale 'Emotional Problems'	7	7	✓	1 - 5	x	00:47	.86-.88	✓	0.0 (HRV) - 31.6 (SWE)	2.13 (0.77) - 2.65 (0.77)
Social-Awareness										
<i>Interpersonal Reactivity Index (IRI; Davis, 1980) - Multidimensional assessment of Empathy:</i>										
Scale 'Fantasy'	7	-	✓	1 - 5	x	01:15	.78-.82	✓	0.0 (SVN) - 35.8 (SWE)	3.10 (0.70) - 3.32 (0.69)
Scale 'Empathic concern'	7	7	✓	1 - 5	x	01:15	.69-.76	-	0.0 (HRV) - 34.7 (SWE)	3.84 (0.48) - 4.07 (0.56)
Scale 'Perspective taking'	7	7	✓	1 - 5	x	00:50	.81-.83	✓	0.0 (SVN) - 42.1 (SWE)	3.48 (0.54) - 3.75 (0.54)
Scale 'Personal distress'	7	-	✓	1 - 5	x	00:50	.75-.84	-	0.0 (SVN/HRV) - 42.1 (SWE)	2.20 (0.63) - 2.78 (0.58)
<i>Vignette: Social Perspective Taking (Diazgranados et al., 2016)</i>	n.a.	3		n.a.		10:00				
Relationship skills										
Teachers' Relational Competence Scale (TRCS; Vidmar & Kerman, 2016)	11	9	✓	1 - 5	x	01:13	.79-.92	✓	0.0 (SVN/HRV) - 66.3 (SWE)	3.86 (0.5) - 4.02 (0.44)
Classroom Climate										
Teacher Evaluation of a Positive Climate in the Classroom (Bear et al., 2016)	8	8	✓	1 - 5		01:22	.90-.94	✓	0.0 (HRV) - 63.2 (SWE)	3.31 (0.58) - 3.58 (0.65)
Orderliness of the Classroom (OOC-T; Sullivan et al., 2012; Sullivan et al., 2014)	11	11	✓	1 - 5		02:03	.87-.88	-	0.0 (HRV) - 68.4 (SWE)	1.94 (0.74) - 2.62 (0.81)
Verbal and physical violence among students (Sullivan et al., 2012; Sullivan et al., 2014)	8	8	✓	1 - 5		01:32	.84-.88	✓	0.0 (SVN/HRV) - 65.3 (SWE)	1.58 (0.61) - 1.84 (0.69)
Verbal and physical violence towards the teacher (Sullivan et al., 2012; Sullivan et al., 2014)	5	-	✓	1 - 5		01:01	.66-.85	-	0.0 (HRV) - 64.2 (SWE)	1.06 (0.19) - 1.29 (0.53)
Inter(trans)cultural competence										
Critical Consciousness Scale (CCS; Diemer, Rapa, Park & Perry, 2017) – Critical Reflection	10	10	✓	1 - 4	x	01:07	.86-.90	✓	0.0 (HRV) - 76.8 (SWE)	2.51 (0.49) - 2.88 (0.64)
Reflexivity (Denson et al., 2017)	3	-	✓	1 - 4			.84-.88	-	3.6 (HRV) - 67.4 (SWE)	3.22 (0.49) - 3.36 (0.43)
Adaptability/Flexibility (Denson et al., 2017)	17	5	✓	1 - 6			.62-.84	-	0.0 (HRV) - 60.0 (SWE)	3.85 (0.45) - 4.18 (0.51)
Professional Beliefs about Diversity Scale (PBDS; Pohan & Aguilar, 2001)	24	-	✓	1 - 4		04:16	.73-.75	-	4.0 (SVN) - 100 (SWE)	2.78 (0.30) - 2.92 (0.32)
<i>Vignette: Intercultural Awareness (Schwarzenthal, 2017)</i>	n.a.	3		n.a.	x	10:00				
<i>Vignette: Intercultural Awareness (own development)</i>	n.a.	3		n.a.		10:00				
Culturally Inclusive Teaching Strategies (CITS; Denson, Ovensden, Wright, Paradies & Priest, 2017)	4	4	✓	1 - 4		00:51	.7-.9	-	0.0 (SVN) - 66.3 (SWE)	2.14 (0.72) - 2.42 (0.84)
Teacher Self-Efficacy Scale for Classroom Diversity (TESCD; Kitsantas, 2012)	17	4	✓	1 - 4		01:53	.82-.88	-	1.3 (SVN) - 70.5 (SWE)	2.93 (0.43) - 2.98 (0.51)
Measures targeting several areas										
Focus group interviews	n.a.	n.a.		n.a.	x					
Other measures										
Job Satisfaction Scale (JSS; OECD, 2013b; OECD, 2014)	10	10	✓	1 - 6		01:53	.83-.88	✓	3.6 (HRV) - 76.8 (SWE)	4.59 (0.79) - 4.85 (0.76)

Note. Instruments printed in black font in the "measures" column are part of the final Field Trial questionnaires. The "pilot" column indicates which of those instruments were tested in the Pilot Study. Instruments or single scales appearing in grey font were excluded after the Pilot Study. In the "range" column, the possible response range for each scale is displayed so that the mean values in the "distribution M (SD)" column can be interpreted in relation to it. The values in the "missings (%)" column refer to individual items within a scale: For each scale, a percentage is given for the item with the lowest number of missings and for the item with the highest number of missings (each in a country comparison). The "EFA" column shows whether the factor solution, i.e. the structure across the school systems was comparable (= "✓") or inconsistent (= "-").

Chapter 4:
The HAND in HAND Field Trials: samples and research questions

Mojca Rožman, Nina Roczen, Svenja Vieluf

4.1. Assessment design

The field trials took place in three school systems: Croatia, Slovenia and Sweden. The student target population was 13- to 14-year-olds or grade 8 students. Similarly, the school staff target population was defined as teachers working in grade eight and other school staff in these schools. In addition, only those teachers of the 8th grade class selected for the participation in the project (regardless of whether the student programme was implemented or not) were invited to participate in the teacher programme. Further, representatives of the school leadership and other school staff (e.g. counsellors, school social workers, school psychologists, school nurses) at the same schools completed a HAND IN HAND programme for school leaders and counsellors.

For the estimation of causal effects of the programmes, the study used an experimental design, with (A) one control group and three experimental groups: (B) student training, (C) training for teachers and training for school leaders/ other school staff, (D) student training and training for teachers and training for school leaders/ other school staff. In all four groups a prior measurement (HAND in HAND assessment) was realized. Consequently, students in classes allocated to experimental groups B and D and teachers as well as school leaders and other school staff in schools allocated to experimental groups C and D were subjected to the HAND in HAND programme, no HAND in HAND programme was implemented in schools that belonged to the control group. Directly after the programme implementation in the three experimental groups a post-measurement took place, and a follow-up measurement 6 months later. The programmes and measurements were implemented in three different school systems (Croatia, Slovenia and Sweden). Hence, we have a 4 (experimental groups) x 3 (points in time) x 3 (school systems) design. The assessment design is presented in Figure 4.1.

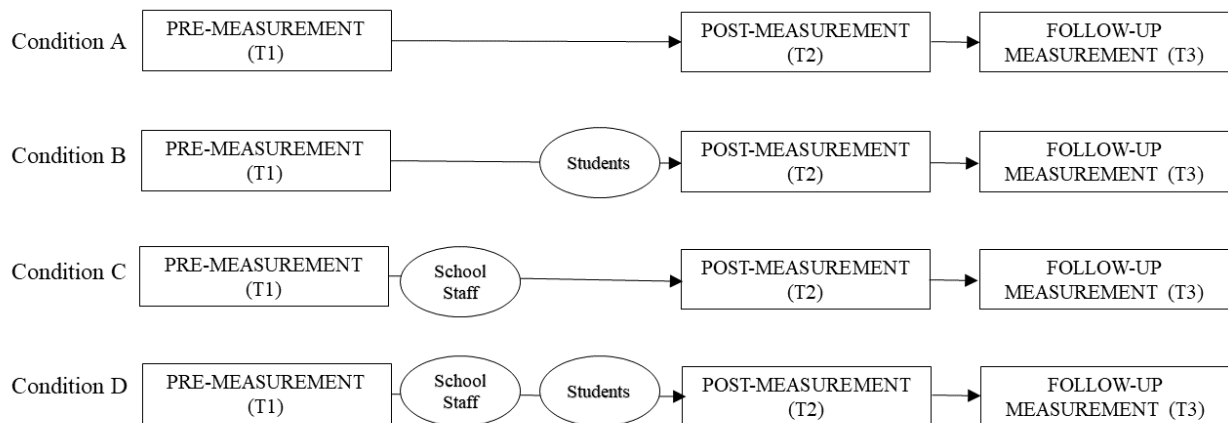


Figure 4.1. HAND in HAND assessment design

The experimental procedure was followed consistently in all school systems conducting the HAND in HAND field trial. Each country made a list of eligible schools meeting the criteria of the target group of “schools with a high percentage of students at risk” (e.g. refugee students and/or other students with a migration background, students from other minorities groups such as Roma students and/or other students with disadvantaged backgrounds). Each country developed a unique sampling plan based on the national context. While the Slovenian team concentrated on schools with students who had recently migrated to Slovenia and needed additional hours of support in the Slovenian language (operationalised by the number of extra hours for Slovenian language lessons offered at the school level), the Croatian team placed its emphasis on schools with significant proportions of Roma children, children from families who had migrated from other ex-Yugoslav school systems (mainly Bosnia and Herzegovina), and schools with other immigrant children. Sweden mainly considered schools containing many students from a disadvantaged socio-economic background and schools with a larger share of immigrant students.

Slovenia provided a sampling frame of eligible schools (which offered 115 or more extra hours in the 1st and 2nd year), from which 14 schools were randomly sampled (12 plus 2 backup schools⁶). In Croatia and Sweden, the national centres

⁶ As two of the initially selected schools refused to participate, two backup schools were included instead.

chose the schools based on the criteria described above. In the next step, the study was presented to all of the selected schools by a researcher from the national HAND in HAND team. They were asked whether they were interested in participating regardless of the experimental condition. After obtaining the consent of the schools, the list of 12 schools was sent to the evaluation team where the schools were randomly allocated to the various (treatment) groups⁷. To select classes within schools the researcher teams in Slovenia and Croatia also provided a list of eligible classes and the evaluation team randomly chose one class within each school before assigning the condition. In Sweden the national researcher team decided which class would participate in the study.

4.2. Description of the samples

Overall, a total of 816 students and 368 members of school staff from 36 schools participated in the HAND in HAND programme. The response rates are presented in Tble 4.1. In Croatia, the majority of participants responded to the questionnaire at all three points in time, T1, T2 and T3. In Slovenia there was a drop of 11 % in response rate from T1 to T2 and 26 % from T2 to T3 in the school staff population. In Sweden, 45 % of school staff who answered the questionnaires in T1 did not respond in T2 and the decline was even larger from T2 to T3 (70 %). Also in the student population in Sweden a decline in response rates was observed (20 % for T2 and 55 % for T3). This is partly due to the drop out of one complete school from the control group after the T1 assessment. Finally, only those that participated at both points (T1 and T2 or T1 and T3) in time were included in the analyses. Due to the significant drop in the response rates for T3 in Sweden the middle term effects (changes between T1 and T3) were not calculated for any of the samples.

⁷ After condition assignment, in Sweden two schools switched roles (one school that was assigned to the control group switched with one school that was assigned to the condition with school staff programmeme only).

Table 4.1

Number of participants at different points in time by country

	School staff			Students		
	Croatia	Slovenia	Sweden	Croatia	Slovenia	Sweden
T1 only	8	17	48	0	3	53
T2 only	6	6	10	2	3	21
T1 and T2	83	128	49	266	265	201
<i>T1 and T2</i>	83	128	49	263	265	201
<i>T1 and T3</i>	72	112	29	264	265	117
<i>T1, T2 and T3</i>	71	111	20	263	265	110
Total	110	151	107	268	271	277

In Table 4.2. we present the number of respondents per condition for each country. We can see that in Sweden the sample sized vary between conditions, whereas the sample sizes in Slovenia and Croatia are balanced.

Table 4.2

Number of school staff members and students participating at all points in time by condition and country

	Croatia	Slovenia	Sweden
Teachers			
Control	7	26	3
Students only	7	18	2
School staff only	8	18	5
Students and school staff	9	21	5
School principals			
Control	2	4	0
Students only	3	2	2

School staff only	3	5	0
Students and school staff	3	2	1
Social workers/school counsellors			
Control	8	4	1
Students only	7	4	1
School staff only	7	2	0
Students and school staff	7	5	0
Total School Staff	71	111	20
Students			
Control	68	66	10
Students only	63	63	14
School staff only	65	62	36
Students and school staff	67	74	50
Total Students	263	265	110

The demographic characteristics of students participating in the study are presented in Table 4.3. In Slovenia and Sweden more girls than boys responded to the student questionnaire. In Croatia, the number of boys and girls in the sample is balanced. In Slovenia most students were 13 years old, in Sweden 14 years and in Croatia 40 % were 13 and 53 % were 14 years old.

Regarding the country of birth for students and their parents, in Slovenia, more than 80 % of the students and their parents were born in the country. In Sweden, the percentage of students and their parents born in the country is slightly lower than in Slovenia but above 75 %. In Croatia 90 % of students report being born in the country, 70 % of students report that their mother and 64 % that their father was born in the country.

The majority of students report that they speak Croatian, Slovenian or Swedish, respectively, at home as can be seen from Table 4.3. About 12 % of students in Slovenia report that in addition to the official national language they also speak another language, which is probably due to the selection of schools that offer

additional hours of Slovenian language, whereas in Croatia and Sweden the percentage is slightly lower (8% and 9%, respectively).

Table 4.3

Demographic characteristics of student samples in Croatia, Slovenia and Sweden

	Croatia		Slovenia		Sweden	
	f	%	f	%	F	%
Gender						
Male	129	(48.1)	117	(43.2)	110	(39.7)
Female	127	(47.4)	151	(55.7)	132	(47.7)
Missing	12	(4.5)	3	(1.1)	35	(12.6)
Age						
12 years		(0.0)	38	(14.0)		(0.0)
13 years	108	(40.3)	215	(79.3)	41	(14.8)
14 years	144	(53.7)	15	(5.5)	198	(71.5)
15 years	4	(1.5)	0	(0.0)	9	(3.2)
16 years	1	(0.4)	0	(0.0)	1	(0.4)
Missing	11	(4.1)	3	(1.1)	28	(10.1)
Country of birth: Student						
Country of test	242	(90.3)	241	(88.9)	214	(77.3)
Other country	11	(4.1)	25	(9.2)	32	(11.6)
Missing	15	(5.6)	5	(1.8)	31	(11.2)
Country of birth: Mother						
Country of test	188	(70.1)	222	(81.9)	210	(75.8)
Other country	69	(25.7)	44	(16.2)	41	(14.8)
Missing	11	(4.1)	5	(1.8)	26	(9.4)
Country of birth: Father						
Country of test	172	(64.2)	221	(81.5)	209	(75.5)

	Croatia		Slovenia		Sweden	
	f	%	f	%	F	%
Other country	80	(29.9)	47	(17.3)	39	(14.1)
Missing	16	(6.0)	3	(1.1)	29	(10.5)
Language spoken at home						
Slovenian/Croatian/Swedish	228	(85.1)	220	(81.2)	210	(75.8)
Other language(s)	4	(1.5)	15	(5.5)	14	(5.1)
Slovenian/Croatian/Swedish and (an)other language(s)	22	(8.2)	33	(12.2)	24	(8.7)
Missing	14	(5.2)	3	(1.1)	29	(10.5)

In Table 4.4. the demographic information on the school staff samples in different school systems is presented. We can observe that in all school systems more women than men responded to the questionnaire. Also most of the respondents from all three school systems were born in the respective country. Croatia has the largest number of participants who were born outside of the country (13 %) compared to Slovenia and Sweden.

Table 4.4

Demographic characteristics of school staff in different school systems

	Croatia		Slovenia		Sweden	
	f	%	f	%	f	%
Gender						
Female	84	(92.3)	122	(84.1)	66	(68.0)
Male	7	(7.7)	23	(15.9)	31	(32.0)
Country of birth						
Country of test	79	(86.8)	139	(95.9)	89	(93.7)
Other country	12	(13.2)	6	(4.1)	6	(6.3)

Table 4.5 presents the age distribution of the participants in different school systems. The average age is the highest in Sweden and the lowest in Croatia but the difference is only marginal.

Table 4.5

Descriptive statistics for school staff's age by country

	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Croatia	42,4	42	8,9	26	63
Slovenia	45,9	46	9,8	25	63
Sweden	47,4	47	10,7	25	69

In the school staff questionnaire a question for principals was inquiring about the percentage of students at school whose heritage language is different from Croatian, Slovene or Swedish, the percentage of students with special needs and the percentage of students that come from socioeconomically disadvantaged homes. The results are presented in Table 4.6. Across school systems, most schools participating in the study have less than 20 % of students whose heritage language is different from Croatian, Slovene or Swedish, respectively. In about half of the participating schools, there are between 10 and 20 % of students with special needs. This holds true for all school systems. In Sweden principals from half of the participating schools report that there are more than 20 % of students coming from a socioeconomically disadvantage homes whereas in Slovenia and Croatia about half of the principals report a percentage between 10 and 20 %.

Table 4.6

Demographic characteristics of schools as reported by school principals

	Croatia		Slovenia		Sweden	
	f	%	f	%	f	%
Percentage of students whose heritage language is different from Croatian, Slovene or Swedish						
Below 10 %	11	(100.0)	4	(40.0)	4	(33.3)
Between 10 and 20 %	0	(0.0)	3	(30.0)	7	(58.3)
More than 20 %	0	(0.0)	3	(30.0)	1	(8.3)

Percentage of students with special needs					
Below 10 %	6	(54.5)	7	(70.0)	1 (8.3)
Between 10 and 20 %	5	(45.5)	4	(40.0)	7 (58.3)
More than 20 %	0	(0.0)	0	(0.0)	4 (33.3)
Percentage of students from socioeconomically disadvantaged homes					
Below 10 %	4	(36.4)	3	(30.0)	1 (8.3)
Between 10 and 20 %	7	(63.6)	5	(50.0)	3 (25.0)
More than 20 %	0	(0.0)	3	(30.0)	6 (50.0)

4.3. Research questions

As described in Chapter 2 in this book, the HAND in HAND project covers different types of evaluation. The following section lists the questions guiding the external evaluation (experimental outcome evaluation and interview-based evaluation) and the internal evaluation (quality assurance and evaluation of the implementation).

4.3.1. *Experimental Outcome Evaluation*

Differences between the control and experimental groups with regard to changes in Social and emotional competencies and intercultural competencies/ diversity awareness and classroom climate were measured at three points in time. First, before the programme implementation (T1), second, shortly after the programme implementation (T2) and, third, 6 months after the programme implementation (T3). Analysing the changes between the pre- (T1) and post-measurement (T2) we can study short-term programme effects. Analysing the changes between pre- (T1) and follow-up measurement (T3) informs about the middle term programme effects. The quantitative outcome-evaluation aimed at answering the following research questions:

- Do we observe an effect of participation in the HAND in HAND student programme on social and/or emotional competencies and/or intercultural competencies/diversity awareness of students?
- Do we observe an effect of participation in the HAND in HAND teacher and school leader/other school staff programmes on social and/or emotional competencies and/or intercultural competencies/diversity awareness of school staff?

- Do we observe an effect of participation in the HAND in HAND teacher and school leader/other school staff programmes on social and/or emotional competencies and/or intercultural competencies/diversity awareness of the students taught by participating teachers?
- Which programme is more effective with regard to causing short-term improvements in the outcomes for students: The student programme or a combination of all three programmes?

In addition to the analysis of statistical effects of the programmes on social and emotional competencies and intercultural competencies/ diversity awareness of students and school staff and the perceived classroom learning environments described in the two questions above, we included in the student and school staff questionnaires questions addressing the participants' perception of training effects and their view on different aspects of the programmes. In this way, also the participants contribute to the evaluation with their feedback. Furthermore, the participants' perspective is given special attention in the interview-based evaluation described in Chapters 7 and 8.

4.3.2. Interview-based Evaluation

Focus group interviews were conducted in each country after the implementation of the programmes. The three participant groups (students, teachers and school leaders/other school staff) were separately interviewed in groups in each participating school. Each respective interview group consisted of two to eight participants in addition to one or two researchers who led the discussion. The main aim of the interviews was to get participants' feedback on the trainings, their evaluation of how easy/difficult it is to implement the training content in school, a description of the atmosphere during the trainings and to get additional information about the context that could influence the programme implementation and possible effects (other events at school during the implementation period). The results of the interview-based evaluation can be used to complement the results of the experimental outcome evaluation and contribute to process evaluation.

4.3.3. Quality Assurance

Quality assurance represents a set of procedures chosen to ensure the desired level of quality in the HAND in HAND project and aims to establish processes and procedures in the workflow that help maintain a good standard of all work in the project. It was guided by the following questions:

- Is the project implemented according to the plan?
- Are the project meetings held in a transparent way, giving each partner possibilities to contribute?
- Are the programmes implemented according to the plan at schools?

4.3.4. *Implementation Check*

The aim of the empirical implementation research in the context of the HAND in HAND programme was to follow the implementation in three school systems over time. The research questions guiding the study were:

- How do trainers perceive the process of implementing the HAND in HAND programme in local schools?
 - What did they perceive as helpful?
 - Which challenges did they report?
 - What did they report having learned in the process?

PART B: RESULTS OF THE SUMMATIVE EVALUATION

Chapter 5:
**Change in social and emotional competencies and intercultural
competencies/diversity awareness: Results from questionnaires⁸**

Mojca Rožman, Nina Roczen, Svenja Vieluf

⁸ Part of this chapter is based on Rožman, Roczen and Vieluf (2020)

5.1. Introduction

The primary aim of the experimental summative outcome evaluation of the HAND in HAND programme was to investigate whether the programme had effects on social and emotional competencies and/or intercultural competencies/diversity awareness of students and/or school staff and/or on the quality of classroom climates in the participating schools as theoretically expected. A review of literature on effects found in previous studies can be found in Kozina, Vidmar and Veldin (2020). In this chapter, we focus on programme effects on self-reports of social competencies, emotional competencies and intercultural competencies/diversity awareness included in the questionnaires and other-reports from the student questionnaire⁹ (the results from the classroom climate scales are presented in Chapter 12 in this report).

In line with the research questions described in Chapter 4, we present results of quantitative analysis of differences between the control and experimental groups with regard to manifest changes in social and emotional competencies and intercultural competencies/diversity awareness between T2 and T1 representing short-term effects, and between T3 and T1 representing middle term effects.

5.2. Methods

First step of the statistical analysis was combing responses to single items in the questionnaire to scale scores. The dimensionality of scales was examined by means of exploratory factor analysis using data from the Pilot Study. The results are presented in Chapter 3 (see also Roczen et al., 2020). We used Cronbach's alpha as a measure of internal consistency. The coefficient for most scales in all three school systems was higher than $\alpha = .70$ (more details can be found in a separate file (Reliabilities_T1_T2_T3.xlsx) available at www.handinhand.si)¹⁰. The scale score for each participant at each point in time was computed as the arithmetic mean of responses to the items of one scale¹¹. A scale value was only computed if responses for at least half the items of a scale were available.

⁹ There were two scales targeting SEI competences in the student questionnaire that were not self-reports but so called other-reports (see Chapter 3 in this book). *Cooperation*: each student got three classmates randomly assigned and assessed those classmates' cooperative ability; *Teachers relational competence*: students were asked to assess the relational competence of their teacher. Other-reports were not administered in Sweden.

¹⁰ The scales with reliabilities between .60 and .70 are marked in the graphs with an *. The scales with reliabilities lower than 0.60 were excluded from analyses.

¹¹ There was one exception to this procedure for the scale *Cooperation*. For this scale, each student was assessed by three classmates. Firstly, a scale score for the peer-assessment from

The HAND in HAND intervention study is characterized by a classroom-level, not student-level, treatment assignment. Hence, the experimental or study unit is a group of subjects (classroom) and so, strictly speaking, the effect of treatment applies at the classroom, rather than at the individual student level. Moreover, the data collected for the HAND in HAND programme has a multilevel structure with students and school staff being nested within classrooms and schools (although, as data were collected from only one classroom per school, the school and classroom levels coincide), and schools being nested within education systems or school systems. This is important to consider in our methodology because students within the same classroom share many unobserved characteristics which might influence our statistical analysis. It was difficult to take the school level into account in the analysis: Given the small sample sizes at the school level, it was impossible to use multi-level modelling [according to Maas & Hox (2005) multilevel modelling requires at least about 20 cases on the highest level, but we only have 12 schools per country]. Therefore, we solely analysed effects at the individual level, taking account of the multilevel-data structure by correcting standard errors for clustering at the school level. Accordingly, we used linear regression analyses of the student and school staff data to predict changes in outcome variables with treatment assignments at the individual level.

To assess the effects of the HAND in HAND programmes, we compared changes in an outcome in the experimental groups to those in the control group. To this end, we calculated the manifest difference score for each participant in a certain outcome variable before and right after treatment (i.e., scale score T2 – scale score T1) and before and 6 months after the treatment (i.e., scale score T3 – scale score T1). This difference was used as the dependent variable in regression analysis. The experimental condition each individual had been subjected to was used as independent variable. The baseline or the comparison group was the control group.

All statistical analyses – descriptive analysis and those used for scale construction – were carried out with IBM SPSS Statistics Version 22 for Windows (IBM Corporation, 2013). We performed all regression analyses using the R statistical programming environment (RStudio Team, 2015) and corrected the standard errors for clustering in all analyses. As the national traditions of social and emotional learning as well as educational policies in this field differ (Štremfel,

each classmate was calculated separately. Then, these three scale scores were averaged for each student.

2020), and the effects are very heterogeneous across school systems, all analyses were performed separately by country.

In the following text the average changes in groups are presented in graphs and only significant differences are pointed out. In addition, for the significant differences we report Cohen's d as a measure of the effects size. A table containing complete information on all the differences between conditions for all school systems can be found in a separate file (RegressionCoefficients.xlsx) available at www.handinhand.si.

The student programme was administered only in experimental groups B and D and the school staff programmes were administered only in experimental groups C and D. Accordingly, effects of the HAND in HAND programme on school staff can be expected mainly in experimental groups C and D. It is, however, possible that a change in student behaviour due to the student programme also leads to a change in the perception and behaviour of teachers even when teachers have not participated in any programme themselves, so that we might additionally observe effects on school staff in group B. At least such effects would not be unintended. Effects on student scales are mainly expected for experimental groups B and D. An (indirect) effect of the school staff trainings on students (in experimental group C) is additionally expected, but only under the condition that teachers implement exercises from the teacher programmes in the classroom and/or that participation in the teacher programme actually has an effect on teachers' everyday pedagogical practice. Finally, there is some indication that whole school approaches might be more effective than programmes addressing only single groups. This suggests that experimental condition D might be even more effective than experimental conditions B and C. To test these complex theoretical expectations, changes in outcomes are presented separately for each of the three experimental groups.

5.3. Short-term programme effects

In the following, programme effects on changes in students' and teachers' self-reported social and emotional competencies as well as intercultural competencies/diversity awareness between T1 and T2 are presented.

5.3.1. Croatia: Students' Self-awareness

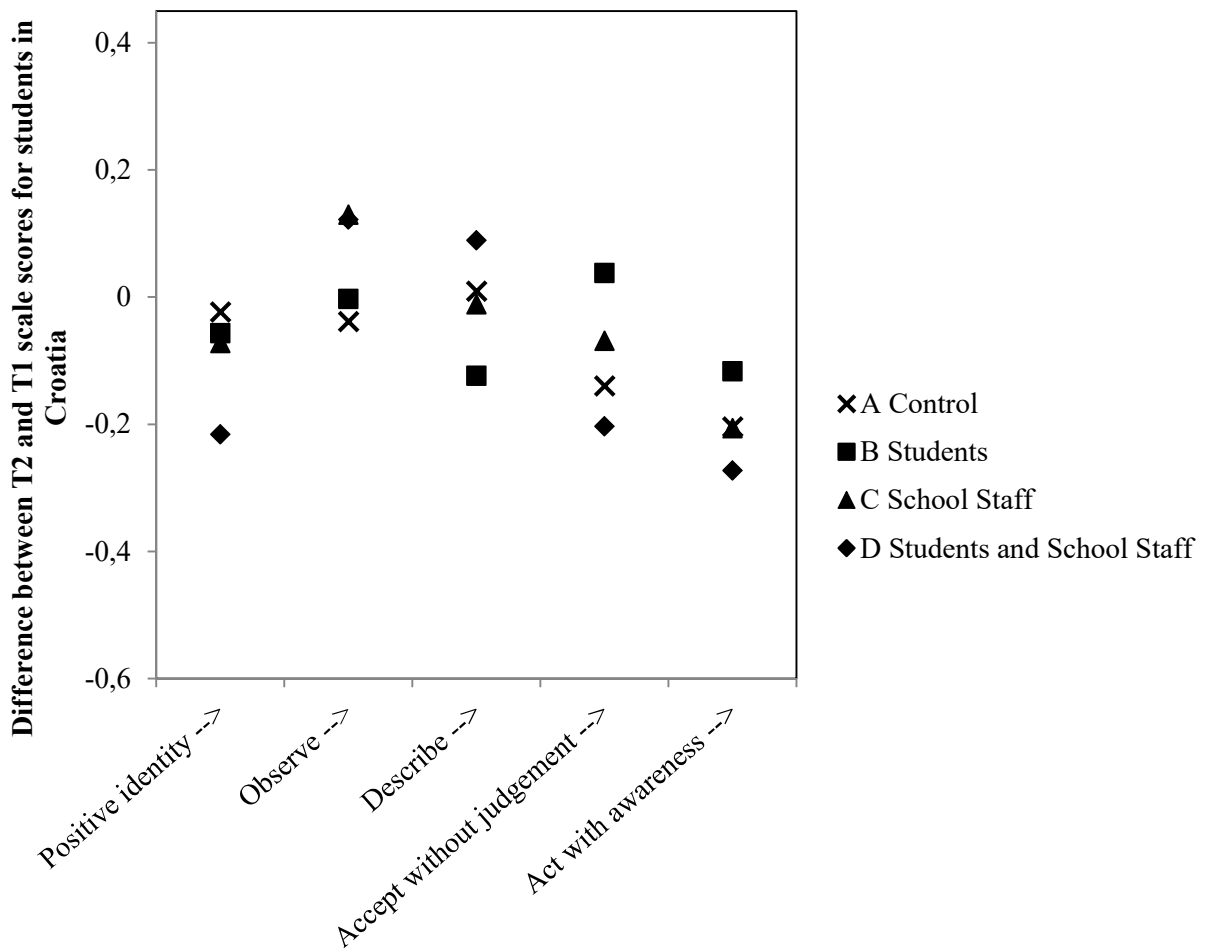


Figure 5.1. Average difference between T2 and T1 per group for the scales assessing the students' self-awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences between groups for scales measuring aspects of self-awareness are shown in Figure 5.1. Three of the 15 differences between the control group and experimental groups were significant. Two of three significant

effects are in the expected direction¹², the third is the opposite to what was expected. This unexpected effect is observed for the scale *positive identity*. The difference in the manifest difference score between the control group and condition D is significant, yet it is only a small effect ($t=-3.61, p=0.000, d=-0.29$). Students from group D report a lower level of *positive identity* at T2 than at T1 while the level in the control group does not change between these points in time.

Effects in the expected direction occur with the scale *observe*. The difference scores for conditions C and D significantly differ from the one in the control group, but both effects are small (C: $t=1.98, p=0.049, d=0.21$; D: $t=2.23, p=0.026, d=0.21$). While the score in *observe* increases for groups C and D between the points in time, it barely changes for the control group, but it also barely changes for group B that participated in the same student programme as classes in experimental group D.

¹² As the programmes aimed to foster social and emotional competencies as well as intercultural competencies/diversity awareness and to improve classroom climates we would expect that specific competencies improve after the programme but also that certain behaviours decrease (for example aggression). In the following text we interpret an increase of a desired trait and a decrease of an undesired trait as an expected effect. The opposite holds true for an unexpected effect.

5.3.2. Croatia: Students' Self-management

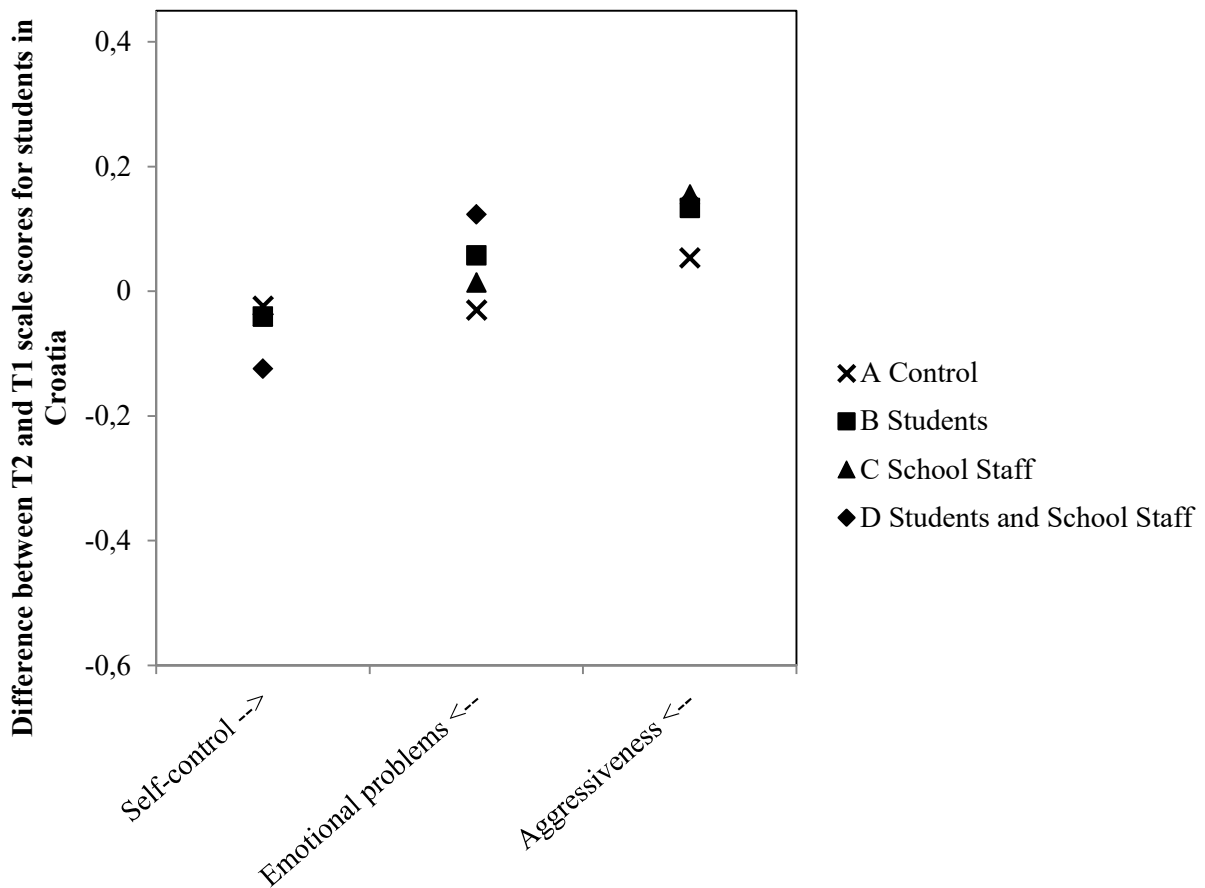


Figure 5.2. Average difference between T2 and T1 per group for the scales assessing the students' self-management in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

We found three significant effects in the unexpected direction for the self-management scales in Croatia. The average differences for the control group and the three experimental groups per scale are shown in *Figure 5.2*. In all three scales, group D significantly differs from the control group – yet these effects are small or even very small (*self-control*: $t=-3.02$, $p=0.003$, $d=-0.18$; *emotional problems*: $t=4.40$, $p=0.000$, $d=0.22$; *aggressiveness*: $t=2.85$, $p=0.005$, $d=0.16$) and groups B and C do not significantly differ from the control group. For students subject to the condition where both, students and school staff, were exposed to the programme we observe a larger average decrease in scale scores for self-control and a bigger increase in emotional problems and aggression compared to the

control group. However, no such effects are observed for group B that also received a student training.

5.3.3. Croatia: Students' Relationship Skills and Social Awareness

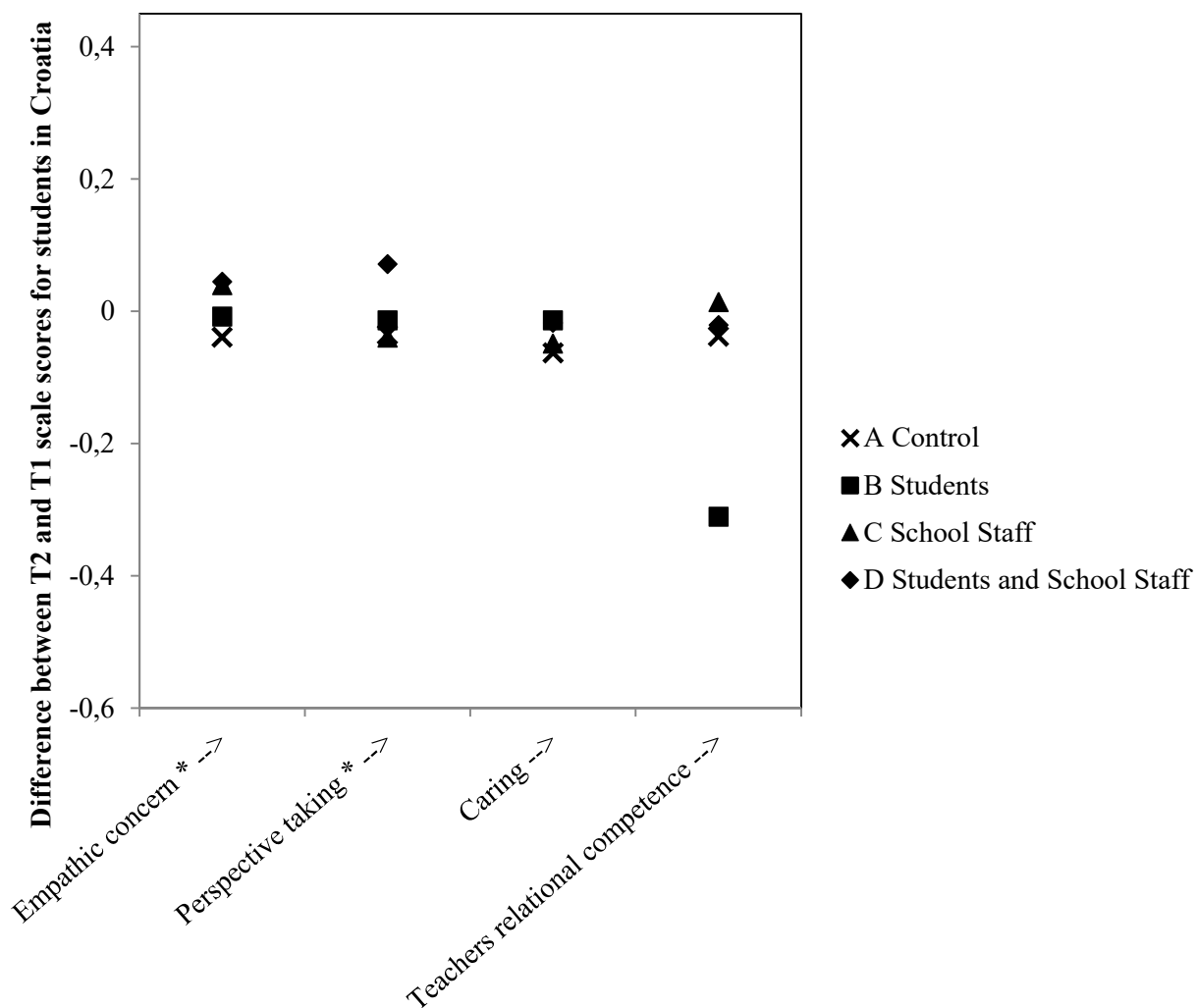


Figure 5.3 Average difference between T2 and T1 per group for the scales assessing the students' relationship skills and social awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was between .60 and .70 in T1 and T2

There is one significant difference between the experimental groups and the control group regarding the size of manifest differences between points in time for the scales that were included to measure relationship skills and social awareness in Croatia. The average differences between the two points in time for

the control group and the three experimental groups and scales are presented in *Figure 5.3*. For *teacher's relational competence* the students from condition B experienced a significantly different manifest change in the self-report measure than the control group, but this effect is only small ($t=-1.98$, $p=0.049$, $d=-0.33$). The effect further points into the unexpected direction: The scale score in condition B decreases while the scale score for the control group does not change much. No such unexpected effect is observed for experimental group D that also participated in the same student programme.

5.3.4. Croatia: Students' Cooperation (Other-reports)

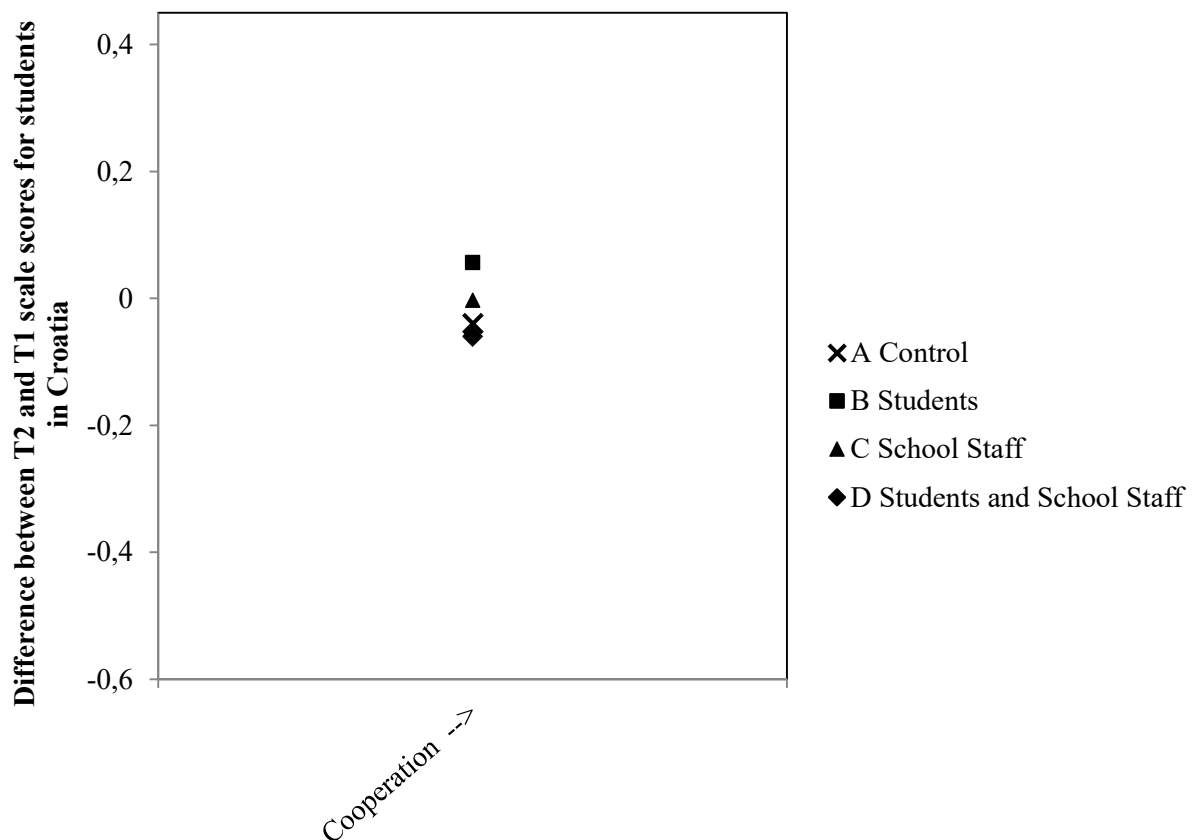


Figure 5.4 Average difference for different conditions between T2 and T1 in students' scale score for cooperation in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Each student's cooperative abilities were assessed by three different classmates that were randomly assigned. This assessment was repeated at each time point by

the same classmates. In Figure 5.4, the average difference (for the averaged scale score) between T2 and T1 is plotted by condition. We find no significant differences in the change score between the control and experimental groups in Croatia.

5.3.5. Croatia: Students' Intercultural Competencies/ Diversity Awareness

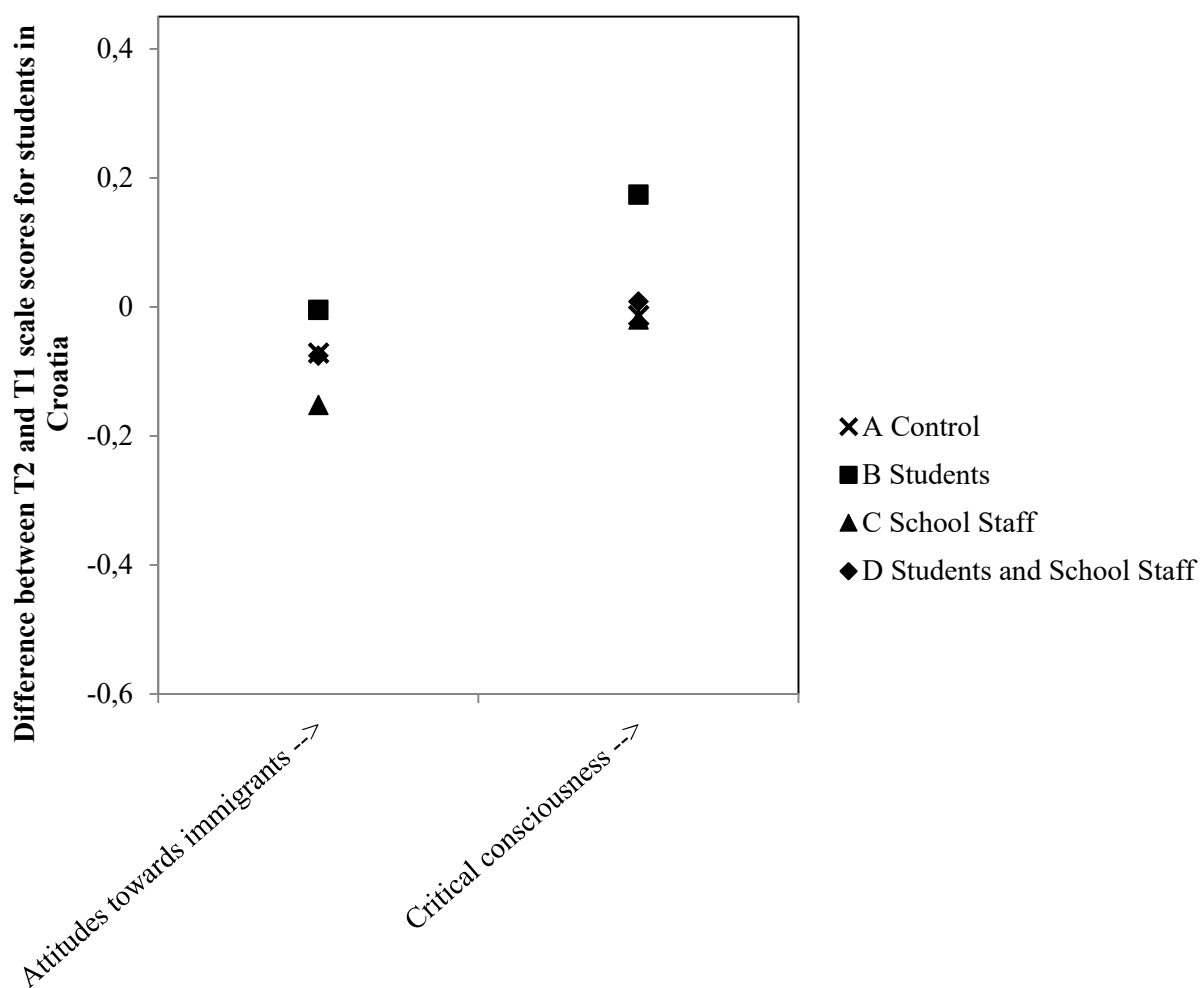


Figure 5.5. Average difference between T2 and T1 per group for the scales assessing the students' intercultural competencies/ diversity awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure intercultural competencies/

diversity awareness are shown in *Figure 5.5*. No significant differences between the control and experimental groups are found for these scales.

5.3.6. Croatia: School Staff's Self-awareness and Self-management

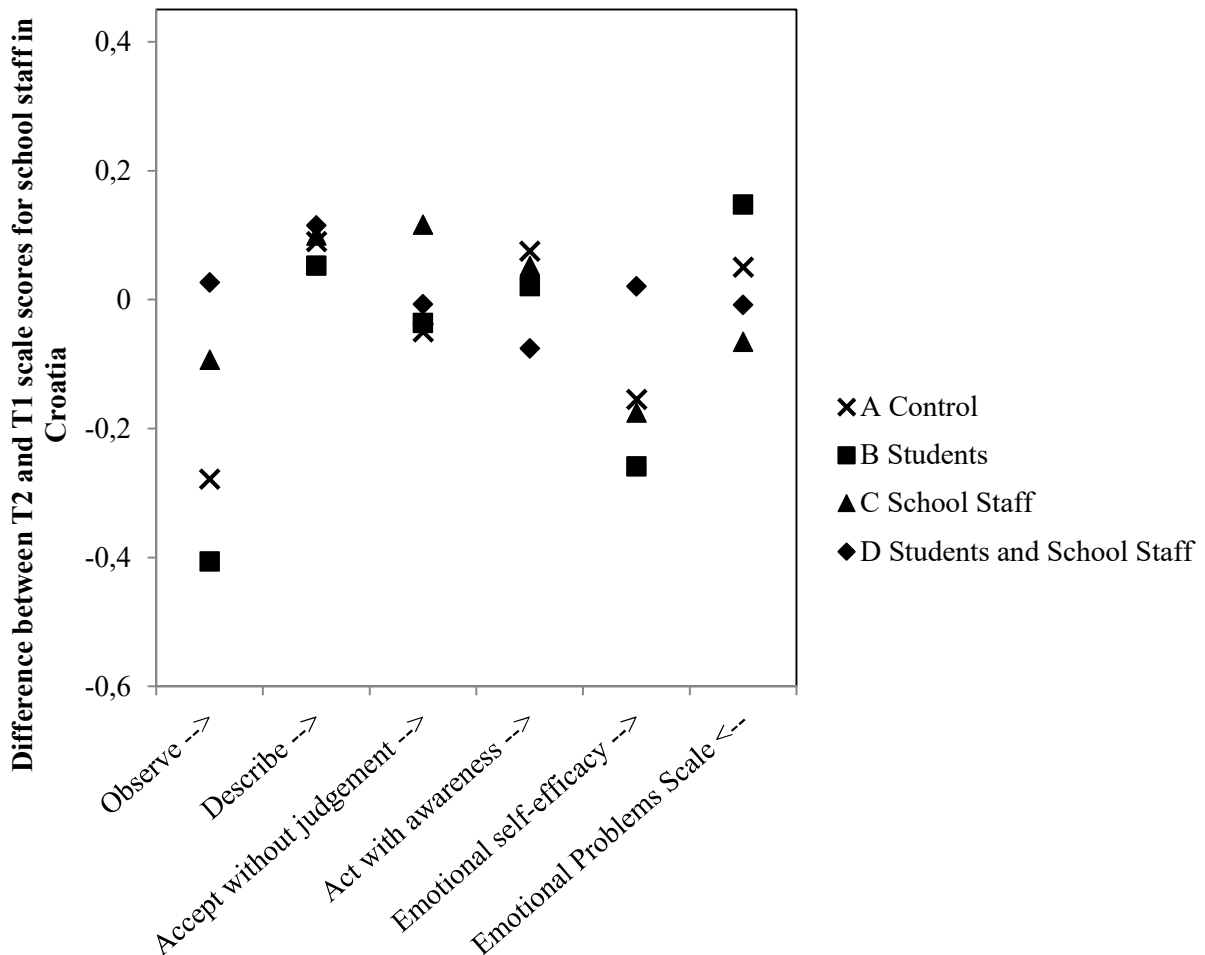


Figure 5.6. Average difference between T2 and T1 per group for the scales assessing the school staff's self-awareness and self-management in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figure 5.6 presents the average differences between the first two measurement-points for different experimental groups regarding self-reported self-awareness and self-management. Only one significant effect can be found for these scales. This effect – that further has the expected direction – is found for the *observe* scale in condition D. The difference between T2 and T1 is, on average, larger in condition D than in the control group, but the effect is rather small ($t=3.07$,

$p=0.003$, $d=0.43$). Although we observe a small decrease between T2 and T1 in the control group, the average scale score in group D does not change. However, not even a small difference is observed between the control group and experimental group C that also received similar school staff trainings as experimental group D.

5.3.7. Croatia: School staff's Relationship Skills and Social Awareness

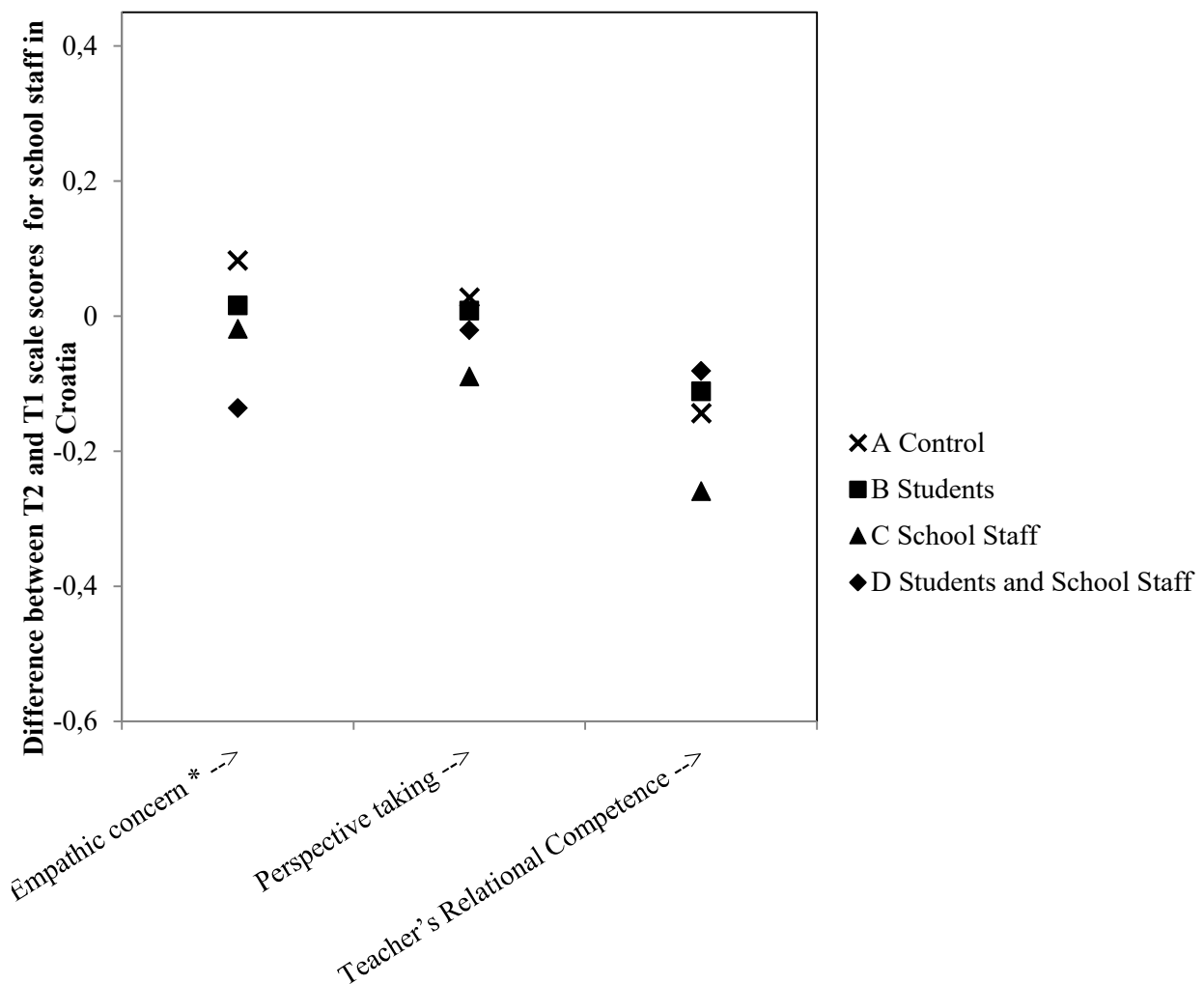


Figure 5.7. Average difference between T2 and T1 per group for the scales assessing the school staff's relationship skills and social awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was between .60 and .70 in T2.

We find no significant effects for the school staffs' relationship skills and social awareness scales. The average differences between the two points in time for the control group and the three experimental groups are presented in *Figure 5.7*.

5.3.8. Croatia: School staff's Intercultural competencies/diversity awareness

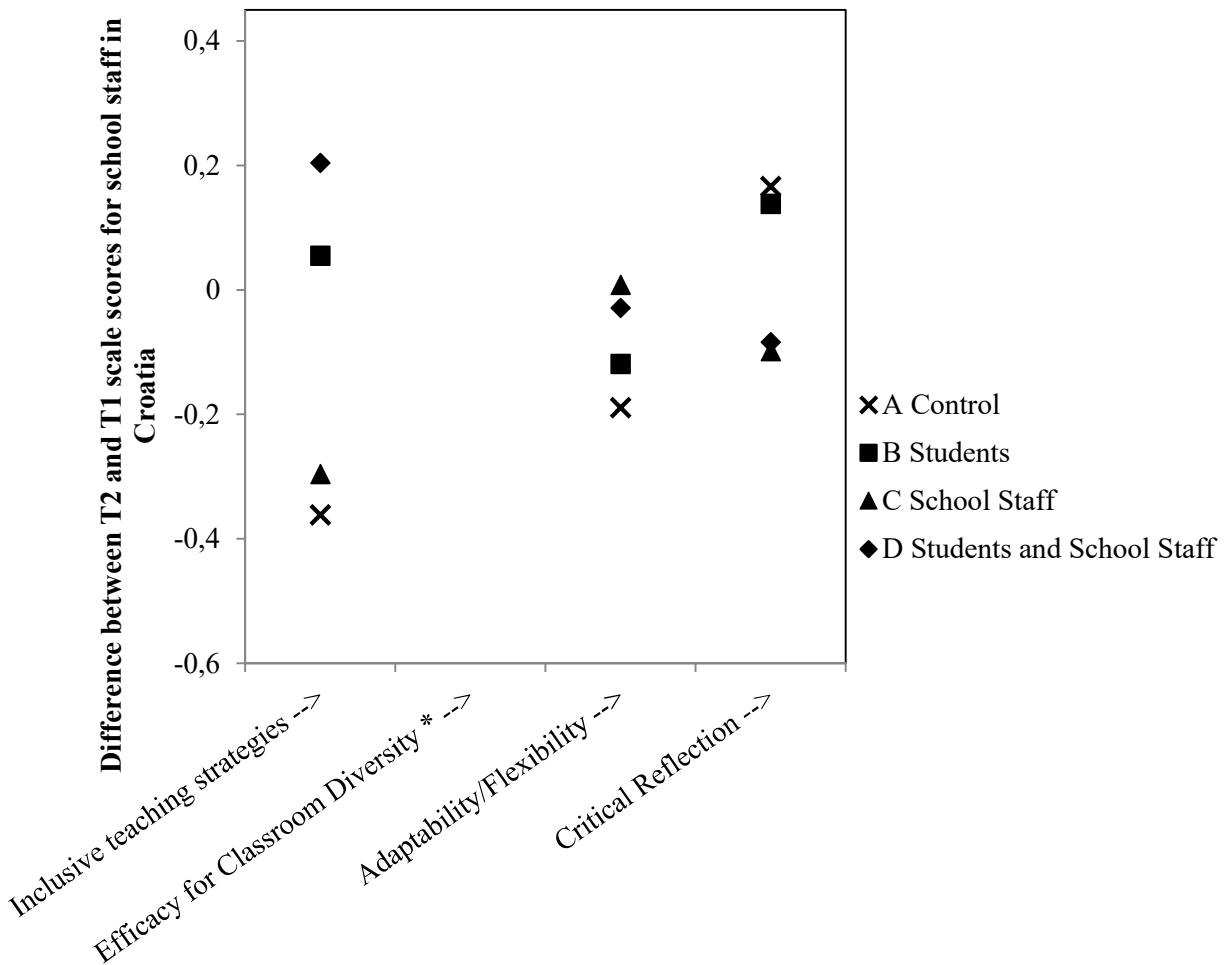


Figure 5.8. Average difference between T2 and T1 per group for the scales assessing the school staff's intercultural competencies/ diversity awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was too low to be reported.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure intercultural competencies/ diversity awareness are presented in *Figure 5.8*. One significant effect in the

unexpected direction is found for *critical reflection*. In group C, there was a decrease in the scale scores whereas in the control group the scores increased from T1 to T2, and this difference is statistically significant but small ($t=-2.64$, $p=0.010$, $d=-0.35$). However, no such difference is observed between group D and the control group, even though group D had the same school staff training as group C.

5.3.9. Slovenia: Students' Self-awareness

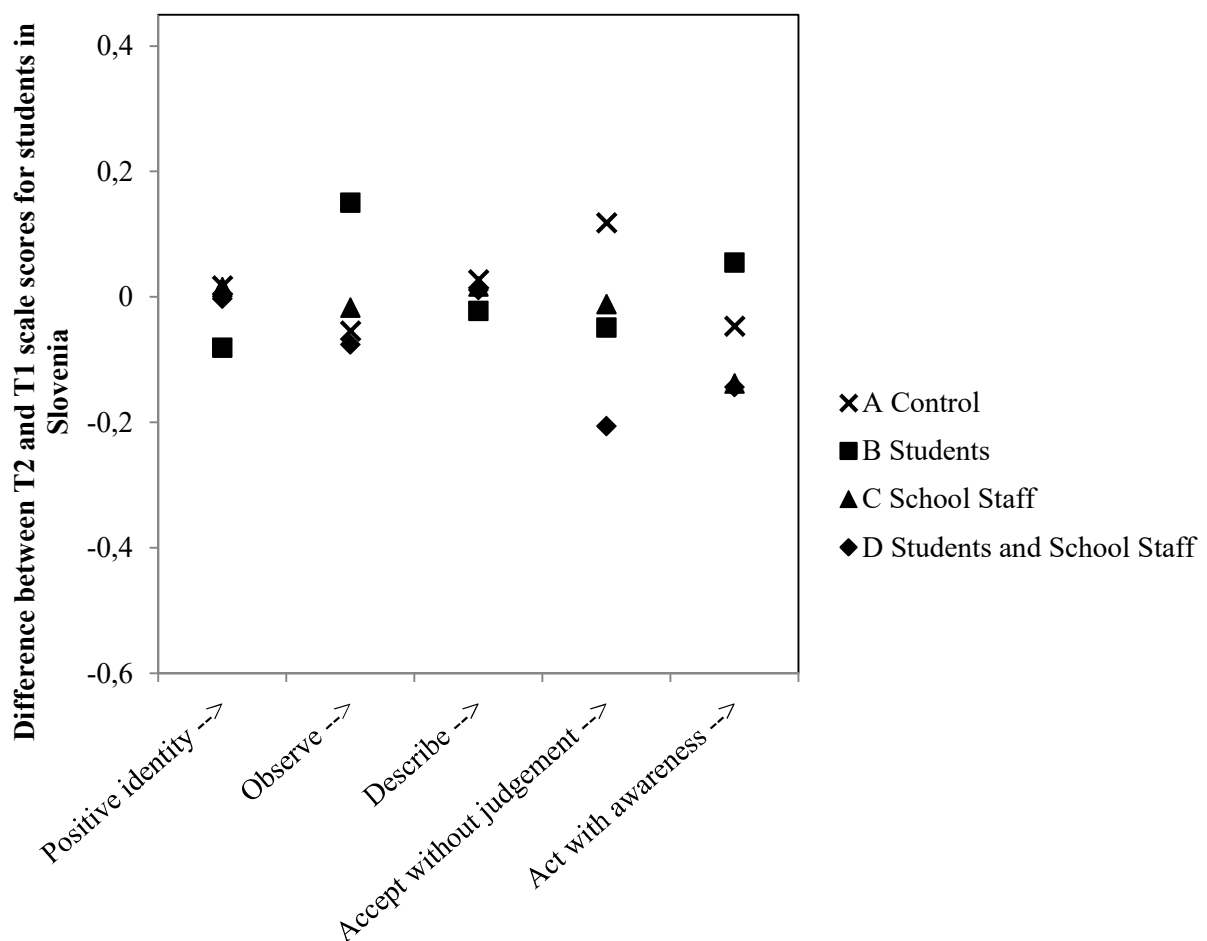


Figure 5.9. Average difference between T2 and T1 per group for the scales assessing the students' self-awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure self-awareness are shown

in *Figure 5.9*. In Slovenia, we can observe one expected – yet small – effect for the *observe* scale. Condition B significantly differs from the control group ($t=3.36, p=0.001, d=0.26$). In the group that completed the student programme, the increase in the scale between T1 and T2 was larger than in the control group. However, no significant difference was observed between condition D and the control group – even though group D had the same student programme as group B.

Moreover, we find unexpected effects in three scales. For *positive identity*, group B significantly differs from the control group – yet this effect is very small ($t=-2.02, p=0.044, d=-0.15$). For *accept without judgement*, a significant effect is found in group D ($t=-2.20, p=0.029, d=-0.37$) and for *act with awareness* in group C ($t=-2.08, p=0.039, d=-0.14$), but the effects are small and very small respectively. The scores in the experimental groups slightly decrease on average, whereas in the control group they do not change for *positive identity* and *act with awareness*, and slightly increase for *accept without judgement*. Again, the effects are not only very small, also the effects of the two groups that had received a student programme are not aligned, so that it is questionable whether the effects can really be attributed to the student programme.

5.3.10. Slovenia: Students' Self-management

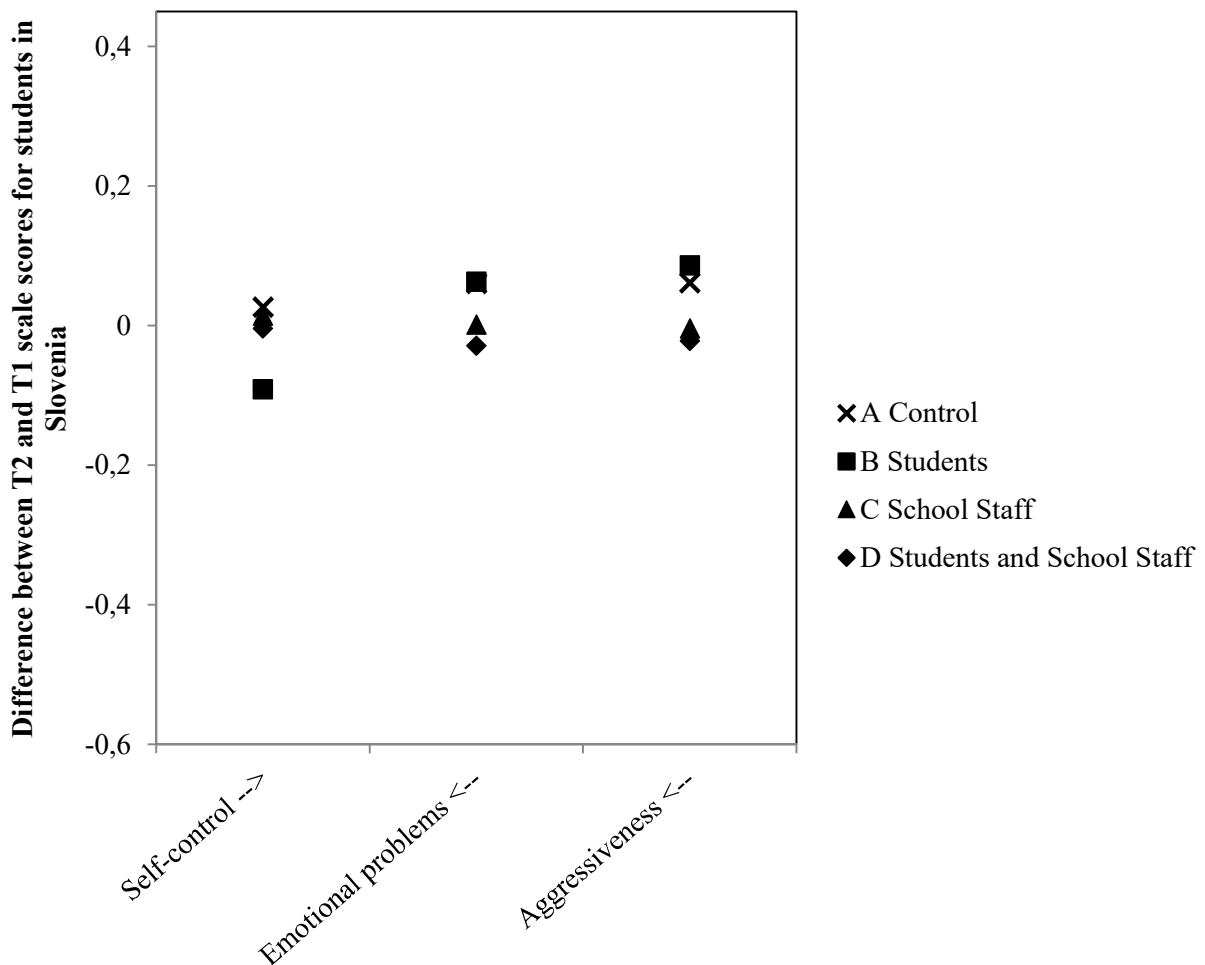


Figure 5.10. Average difference between T2 and T1 per group for the scales assessing the students' self-management in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

There are two significant effects for the scales targeting students' self-reported self-management competencies in the student questionnaire in Slovenia, one in the expected and the other in the unexpected direction. The average differences between the two points in time for the control group and the three experimental groups and scales are presented in Figure 5.10.

The change in *aggressiveness* goes in the expected direction: In condition D, the change between T1 and T2 is larger than in the control group, but the effect is very small ($t=-2.07, p=0.040, d=-0.15$). On average, the scale score in the control

group increased while in condition D it decreased. However, no such difference is observed between condition B and the control group.

The second effect is observed in the unexpected direction for the *self-control* scale. The change in the scale is significantly larger in group B than in the control group ($t=-4.59$, $p=0.000$, $d=-0.22$). The scores in group B on average decrease whereas they do not change much in the control group. Yet, this effect is small and it is not observed for group D in which the same student programme had been implemented as in group B.

5.3.11. Slovenia: Students' Relationship Skills Social Awareness

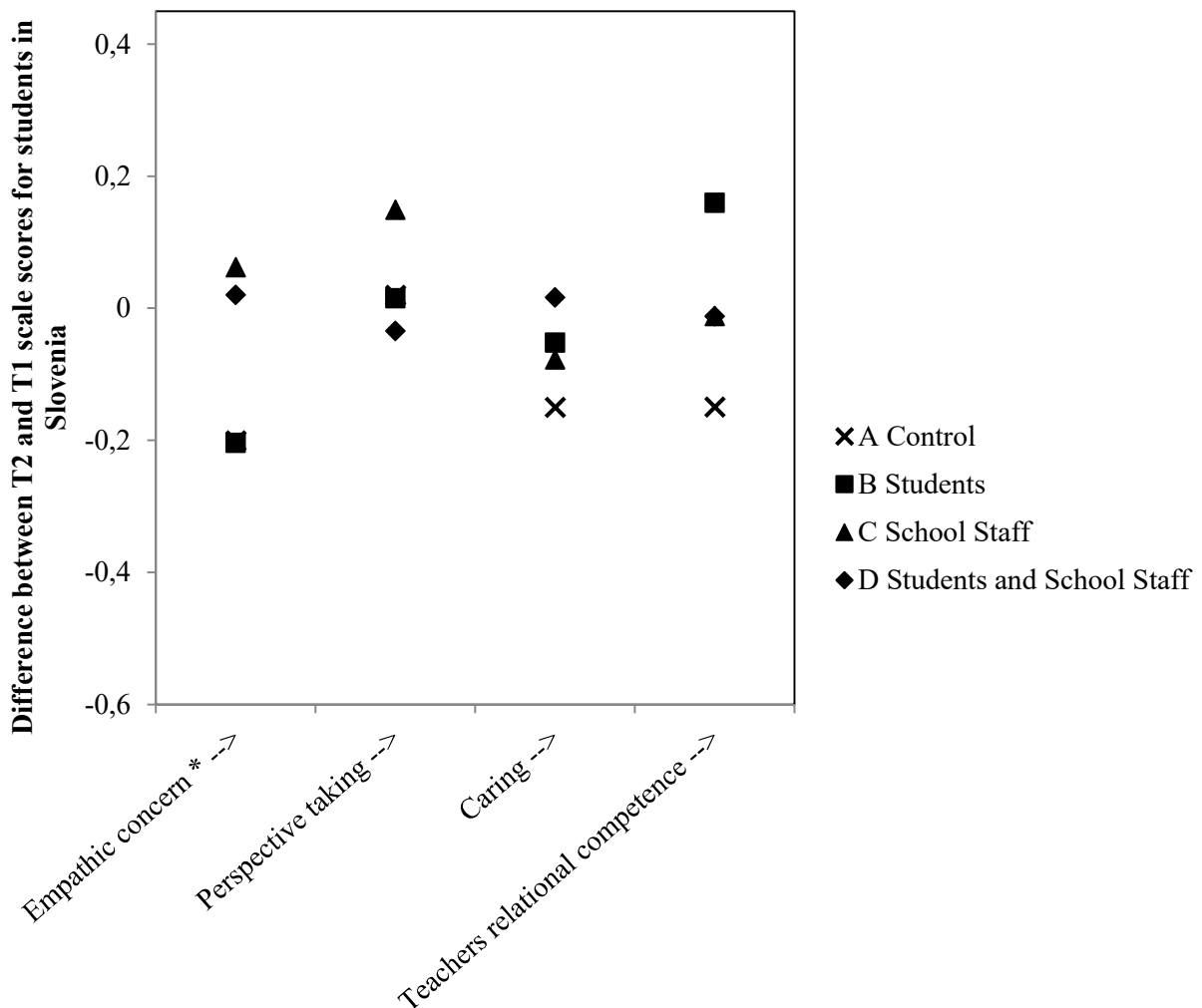


Figure 5.11. Average difference between T2 and T1 per group for the scales assessing the students' relationship skills and social awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was between .60 and .70 in T1 and T2.

Figure 5.11 presents the average differences between the two points in time for the control group and the three experimental groups and scales that measure students' relationship skills and social awareness. We find an effect in the expected direction for *empathic concern*. The changes for groups C and D are significantly larger than in the control group (C: $t=7.79$, $p=0.000$, $d=0.37$; D: $t=3.30$, $p=0.001$, $d=0.32$). While the score in the control group decreases (and also in group B that had received the same student training as group D), it stays on a similar level or slightly increases in groups C and D. Yet, the effect sizes are small. Another effect in the expected direction is found for perspective taking in group C. While the score in group C increased it did not change much in other groups. The difference to the control group is significant but small ($t=3.36$, $p=0.001$, $d=0.20$).

The change in *caring* also goes in the expected direction: In condition D, the change between T1 and T2 is larger than in the control group ($t=2.38$, $p=0.018$, $d=0.27$). On average, the scale score in the control group increased while in condition D it decreased. However, the effect is small and no such difference is observed between condition B and the control group.

5.3.12. Slovenia: Students' Cooperation

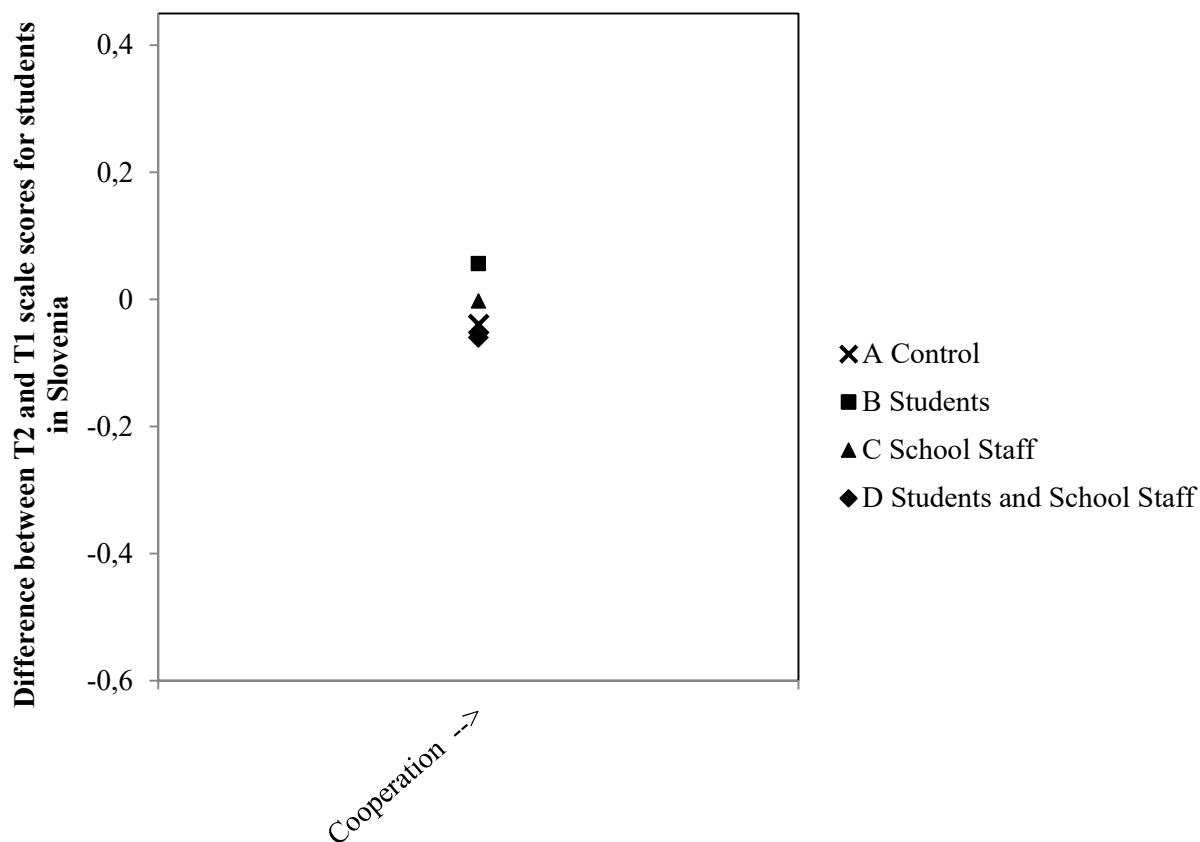


Figure 5.12 Average difference for different conditions between T2 and T1 in students' scale score for cooperation in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences for the students' cooperative abilities between the points in time for different conditions in Slovenia are presented in Figure 5.12. We find no significant effect for *cooperation*.

5.3.13. Slovenia: Students' Intercultural Competencies/ Diversity Awareness

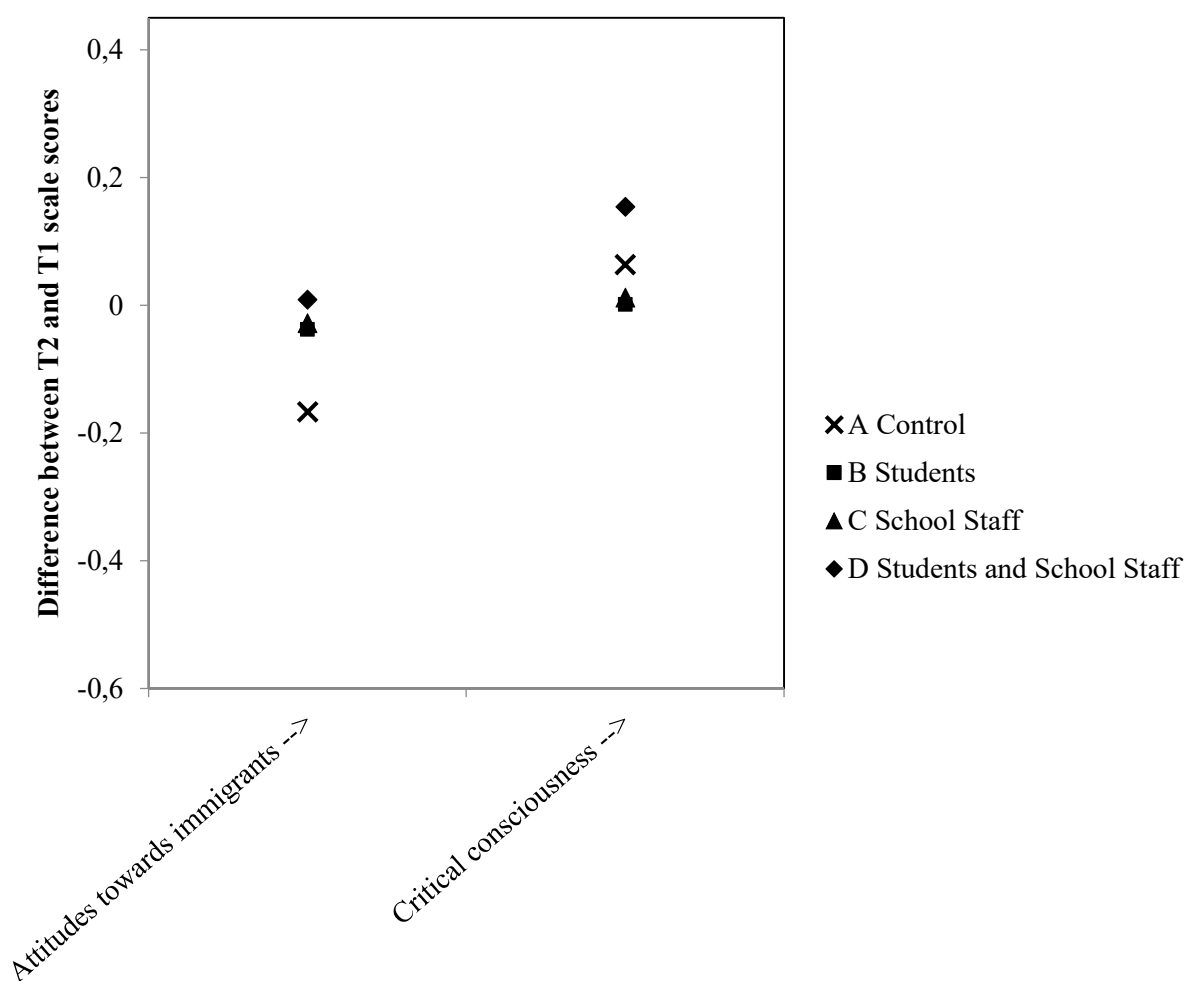


Figure 5.13. Average difference between T2 and T1 per group for the scales assessing the students' intercultural competencies/ diversity awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure students' intercultural competencies/ diversity awareness are presented in Figure 5.13. We found a significant effect in the expected direction for *attitudes towards immigrants*. The difference between the two points in time is significantly larger in groups C and D than in the control group (C: $t=2.40$, $p=0.017$, $d=0.22$; D: $t=3.70$, $p=0.000$, $d=0.27$). While the level of positive attitudes towards immigrants falls from T1 to T2 in the control group, it changes only slightly in experimental groups C and D.

However, the effects are small and no such effect is observed for experimental group B that participated in the same student programme as experimental group D.

5.3.14. Slovenia: School Staff's Self-awareness and Self-management

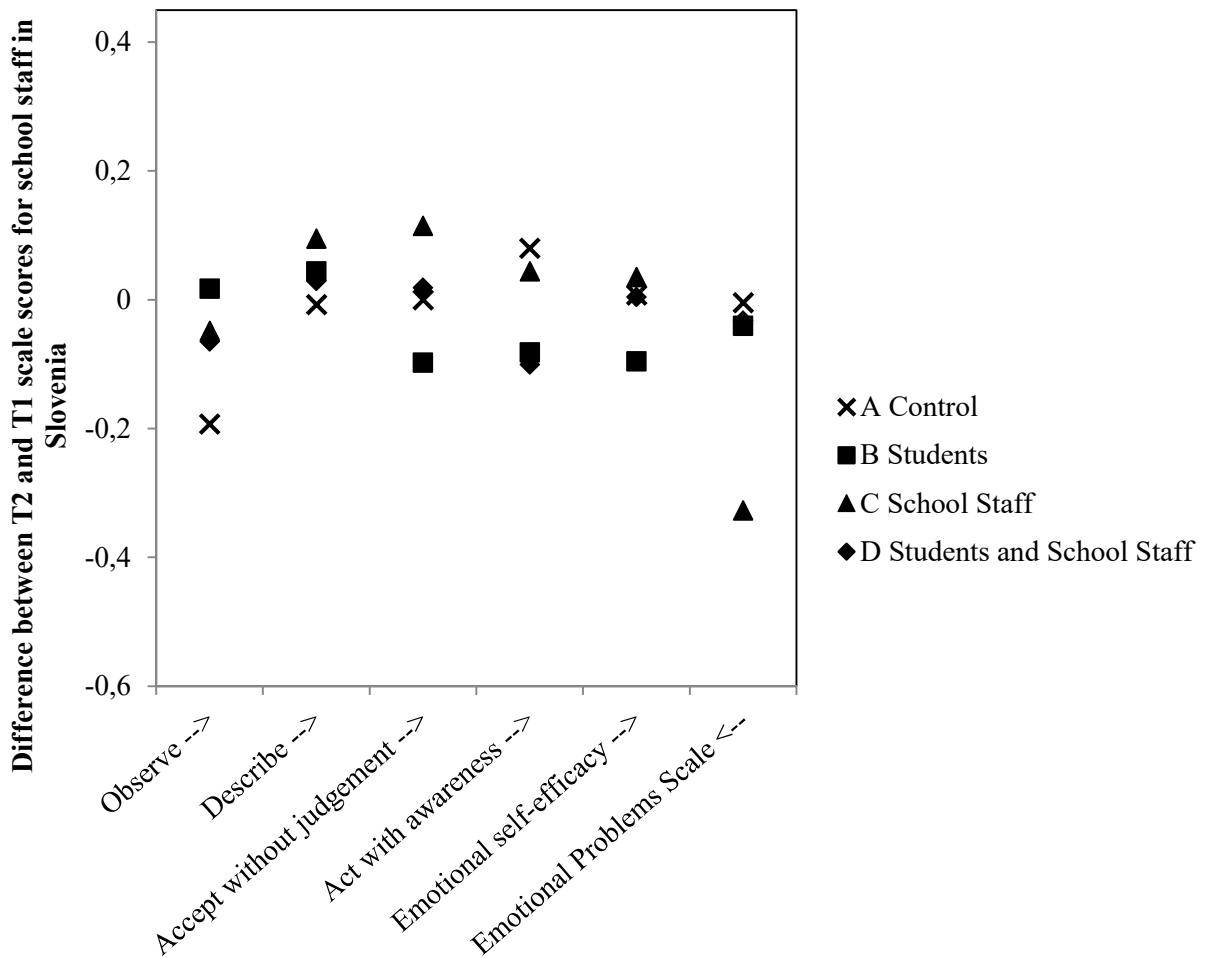


Figure 5.14. Average difference between T2 and T1 per group for the scales assessing self-awareness of the school staff in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure school staff's self-reported self-awareness and self-management are presented in Figure 5.14. One effect in the expected direction is seen with *emotional problems*. School staff from group C show a decrease in emotional problems and this change is significantly different

from that observed in the control group ($t=-2.27, p=0.026, d=-0.46$). However, the effect is rather small and no such difference to the control group is observed for group D.

We find an unexpected significant effect for the *observe* scale. In condition B, the difference between the two points in time is significantly bigger than in the control group ($t=2.87, p=0.005, d=0.28$). While, on average, the scale score does not change in group B, it decreases in the control group – even though no school staff trainings had taken place in neither of the groups. Yet, this effect is only small.

One effect in the unexpected direction is established for the *act with awareness* scale. While in the control group the self-report score increases between the two points in time, it decreases in group D (and also in group B), but not in group C. The differences between groups B (school staff at schools where only a student programme had been implemented) and D (school staff at schools where all three programmes had been implemented) and the control group is significant (B: $t=-2.52, p=0.013, d=-0.27$; D: $t=-2.49, p=0.014, d=-0.27$). However, these effects are small.

5.3.15. Slovenia: School Staff's Relationship Skills and Social Awareness

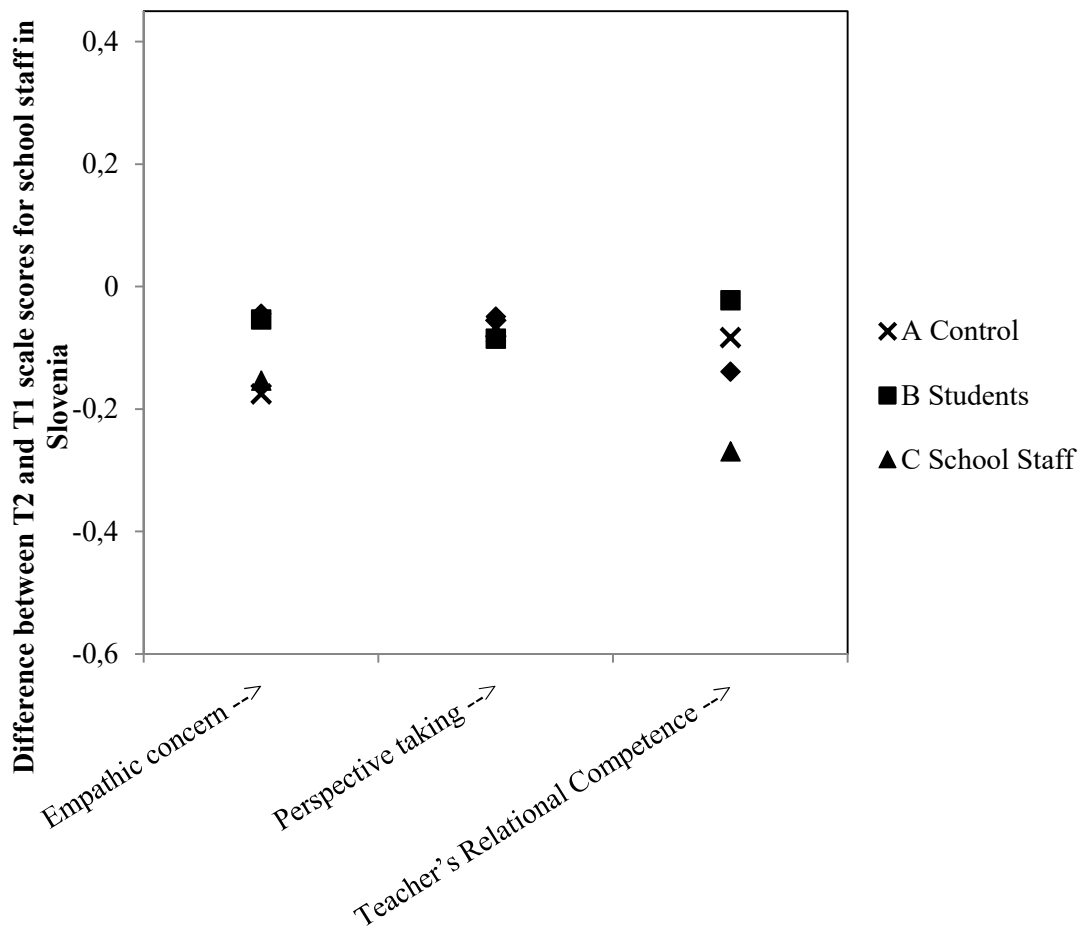


Figure 5.15. Average difference between T2 and T1 per group for the scales assessing the school staff's relationship skills and social awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figure 5.15 shows the average differences between the two points in time for the control group and the three experimental groups and scales that measure school staff's social awareness. An effect in the unexpected direction can be observed for *teachers' relational competence* for group C. Perceived relational competence decreased in group C significantly more than in the control group ($t=-2.22$, $p=0.029$, $d= -0.30$). However, the effect is small and no such effect is observed for group D where the same school staff programmes had been implemented in addition to the student programme.

One effect in the expected direction is found for *empathic concern*. The difference is significant for groups B (even though no staff trainings had taken place for this

group) and D (where school staff participated in trainings) in comparison to the control group (B: $t=2.05$, $p=0.042$, $d=0.19$, D: $t=3.55$, $p=0.001$, $d=0.21$). While the scale score decreases in all groups from T1 to T2, it decreases the least in groups B and D. However, these effects are small and no such effect is observed for group C – even though school staff had in this group had participated in the same training as school staff in group D.

5.3.16. Slovenia: School staff's Intercultural Competencies/Diversity Awareness

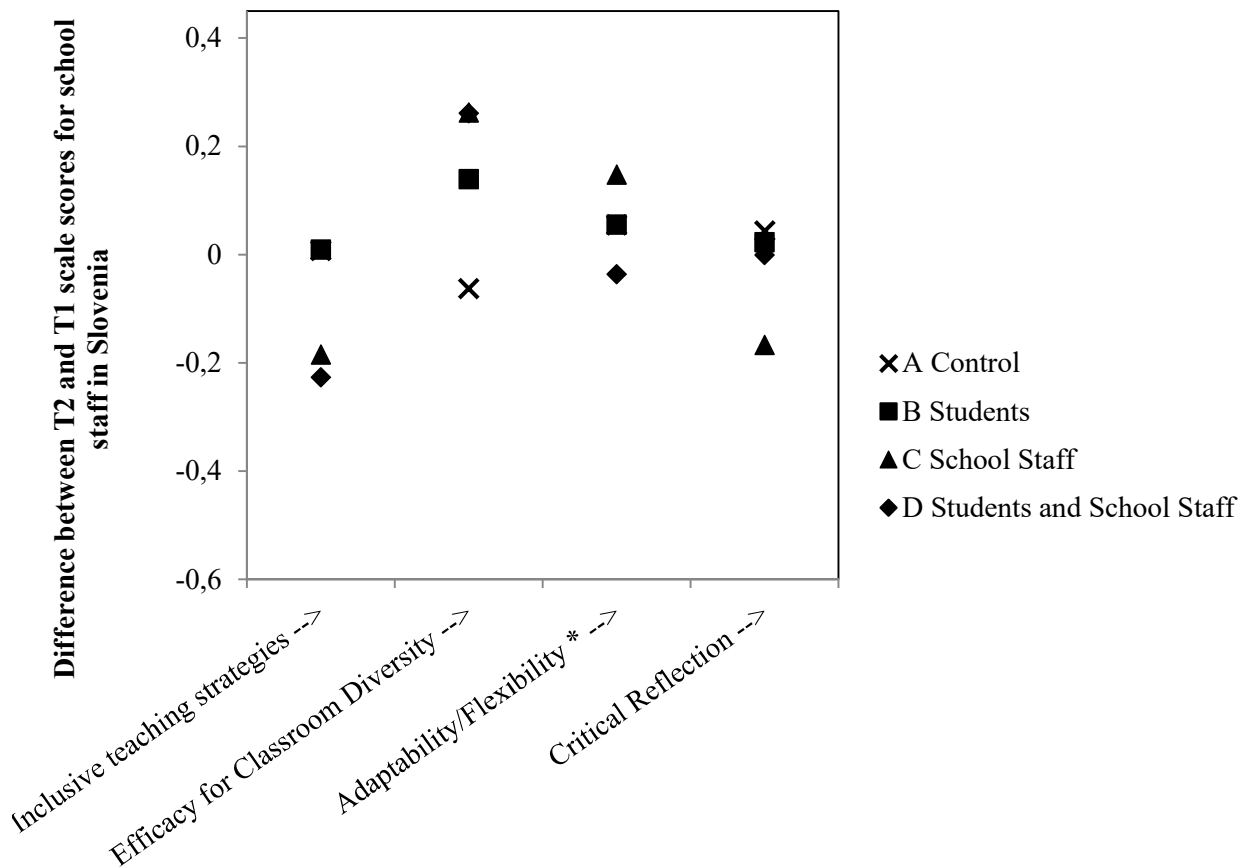


Figure 5.16. Average difference between T2 and T1 per group for the scales assessing the school staff's intercultural competencies/ diversity awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was between .60 and .70 in T1.

Figure 5.16 presents the average differences between the two points in time for the control group and the three experimental groups and scales that measure intercultural competencies/ diversity awareness. On the *efficacy for classroom diversity* scale we find an effect in the expected direction for group D. While in the control group, the self-reported efficacy decreases slightly from T1 to T2, it increases in all other experimental conditions. The difference between group D and the control group is significant ($t=4.11, p=0.000, d=0.45$). However, the effect is rather small and no such effect is observed for group C even though school staff in this group had the same training as school staff in group D.

5.3.17. Sweden: Students' Self-awareness

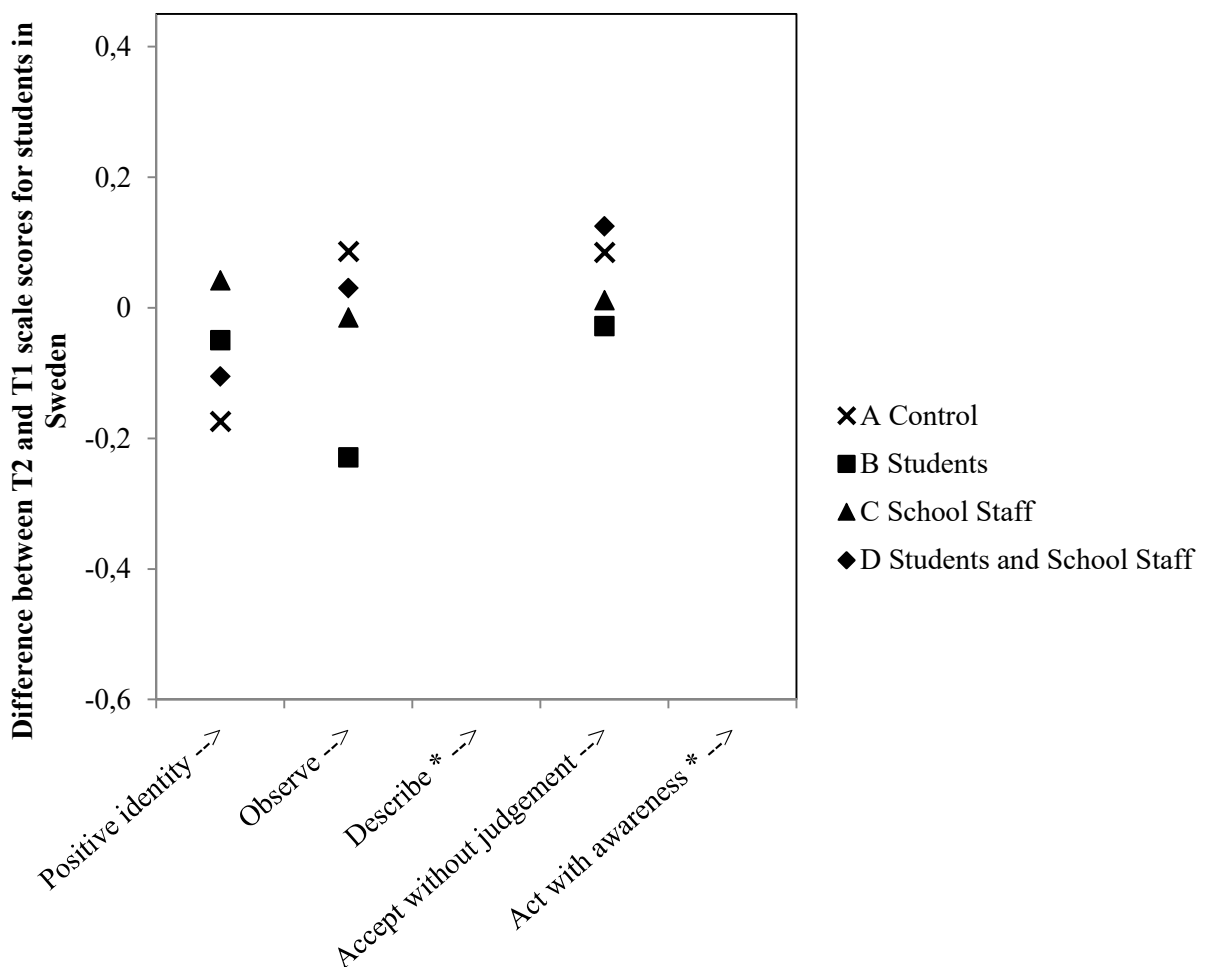


Figure 5.17. Average difference between T2 and T1 per group for the scales assessing the students' self-awareness in Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

* Reliability of this scale was too low to be reported.

For the scales measuring students' self-awareness in Sweden, we find one effect in the expected direction and one in the unexpected direction. In *Figure 5.17*, the average differences between the two points in time are presented. We find significant effects in the unexpected direction for *observe* for group B, compared to the control group (*observe*: $t=-4.17$, $p=0.000$, $d=-0.32$). While the scores in the control group increased for this scale, a decrease was observed in group B. Yet, the effect size is small and no such effect was observed for group D that had been subjected to the same student programme.

The effects of the *positive identity* scale points in the expected direction. While for students in condition C the average scale scores rise from T1 to T2, they fall in the control group. The difference between condition C and the control group is significant (*positive identity*: $t=1.99$, $p=0.048$, $d=0.31$). However, the effect is small and no such effect was observed for condition B or D – even though students in this groups participated in a student programme and those in group C did not.

5.3.18. Sweden: Students' Self-management

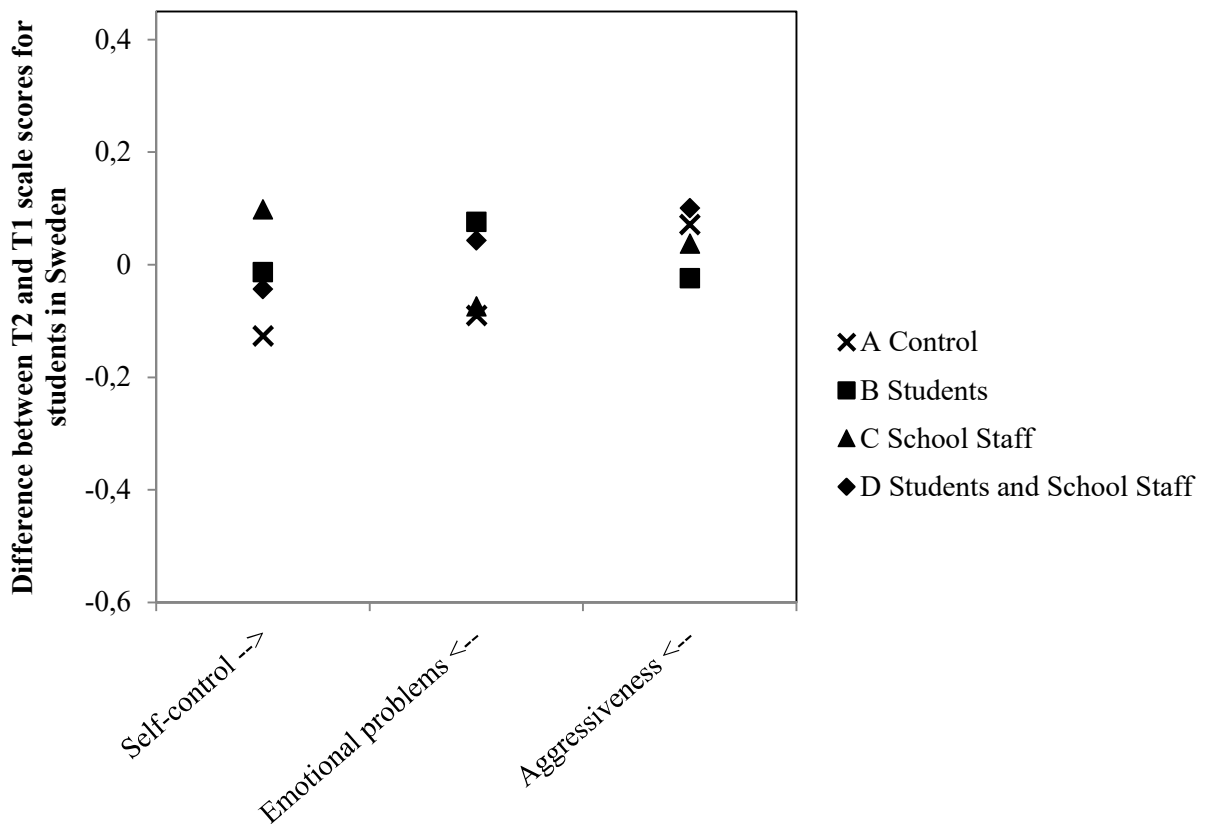


Figure 5.18. Average difference between T2 and T1 per group for the scales assessing the students' self-management in Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figure 5.18 presents the average differences between the two points in time for the control group and the three experimental groups and scales that measure students' self-management. The change in *self-control* between all experimental groups and the control group differs significantly (B: $t=2.78, p=0.006, d=0.19$; C: $t=4.07, p=0.000, d=0.37$; D: $t=2.58, p=0.01, d=0.11$). While the average score in the experimental groups only changes marginally or increases from T1 to T2, it decreases in the control group. However, the effect sizes for groups B and D are very small and that for group C is small.

The change between T1 and T2 for the *emotional problems* scale is also significant, but points in the unexpected direction. For students in group B, we find an increase in emotional problems in T2 compared to T1 whereas the scale score for students in the control group decreases – yet the effect size is small

($t=4.93$, $p=0.000$, $d=0.24$). For students in group D the results also suggest an increase, but the difference between group D and the control group is not significant.

5.3.19. Sweden: Students' Relationship Skills and Social Awareness

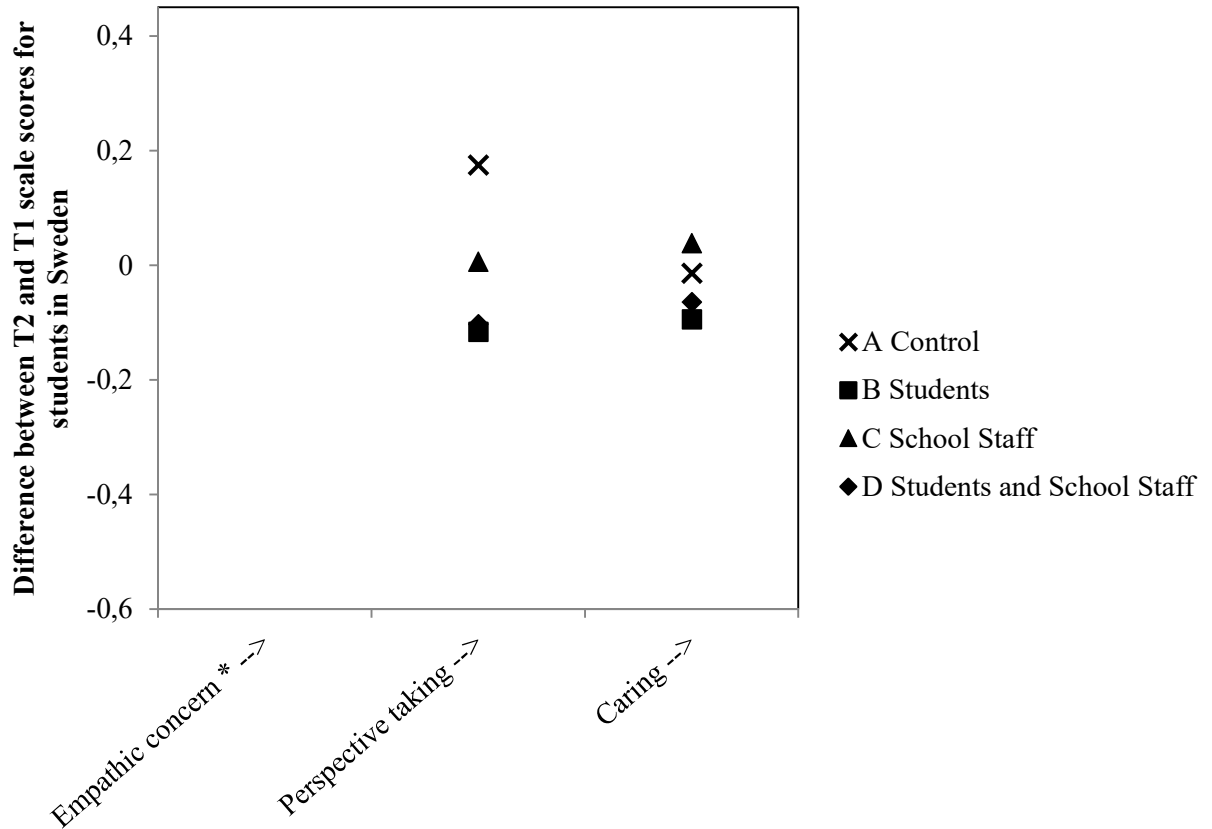


Figure 5.19. Average difference between T2 and T1 per group for the scales assessing the students' relationship skills and social awareness in Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
* Reliability of this scale was too low to be reported.

We find no significant effects for the scales measuring students' relationship skills and social awareness. The average differences between the two points in time for the control group and the three experimental groups and scales are shown in Figure 5.19.

5.3.20. *Sweden: Students' Intercultural Competencies/ Diversity Awareness*

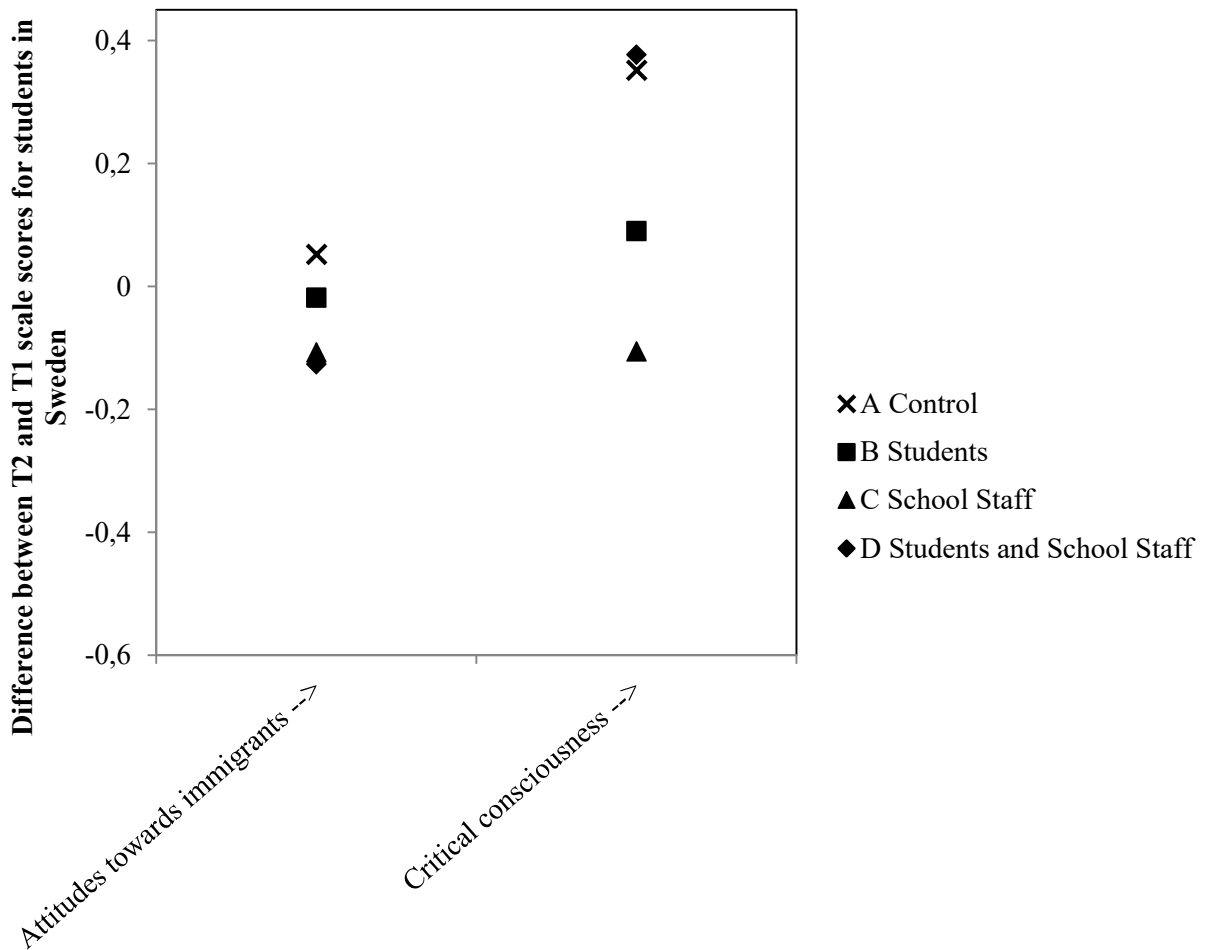


Figure 5.20. Average difference between T2 and T1 per group for the scales assessing students' intercultural competencies/ diversity awareness in Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figure 5.20 presents the average differences between the two points in time for the control group and the three experimental groups and scales that measure intercultural competencies/ diversity awareness. The effect for the *attitudes towards immigrant* scale goes in the unexpected direction. While, on average, the scale scores rise slightly in the control group from T1 to T2, they fall in groups C and D. The difference is significant compared to the control group, but effect sizes are small (C: $t=-2.04, p=0.043, d=-0.22$; D: $t=-2.10, p=0.037, d=-0.26$). However, no such difference to the control group is observed for the other schools that participated in a student training and that belong to group B.

5.3.21. Sweden: School staff's Self-awareness and Self-management

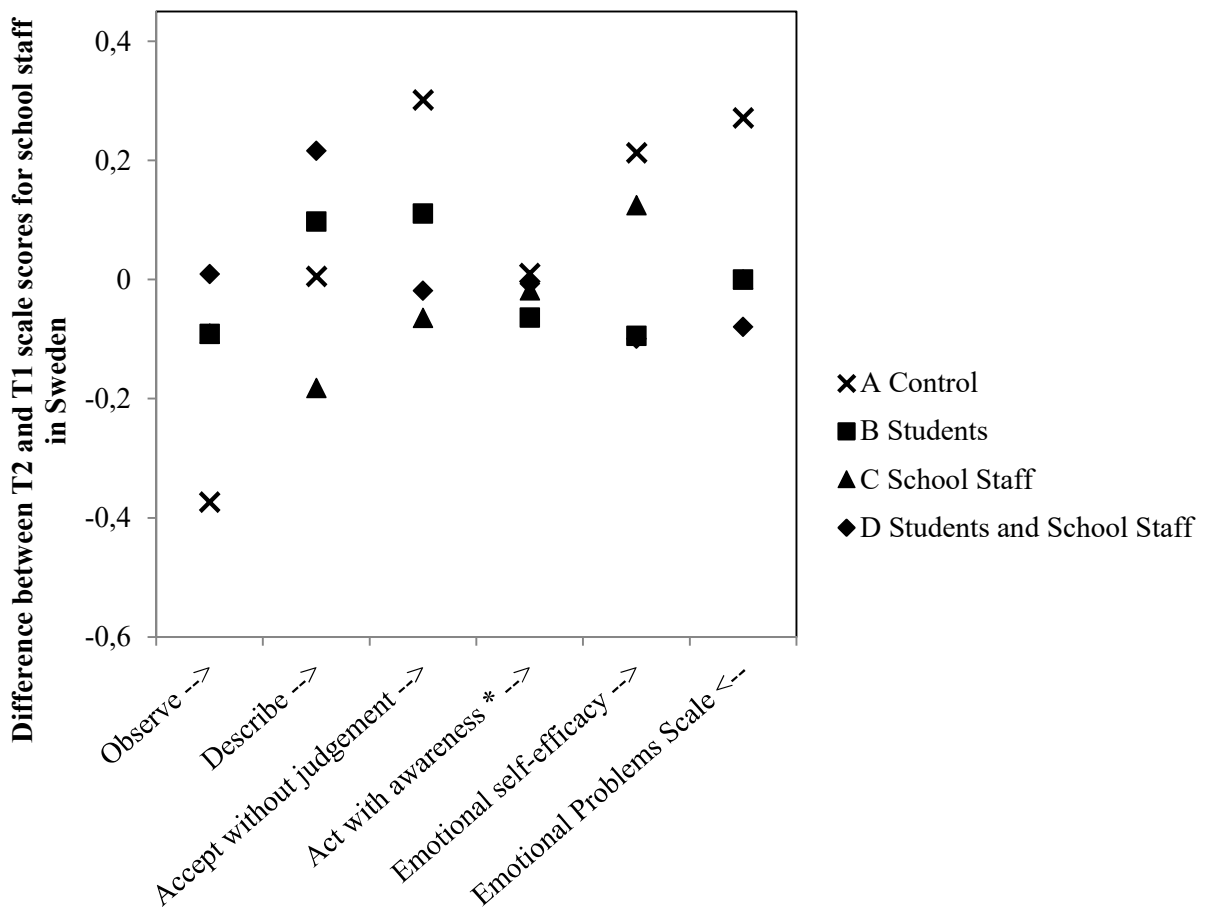


Figure 5.21. Average difference between T2 and T1 per group for the scales assessing the school staff's self-awareness in Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was between .60 and .70 in T2.

In the scales measuring school staff's self-awareness and self-management in Sweden, one effect in the expected and two in the unexpected direction are found. The average differences between the two points in time for the control group and the three experimental groups and scales are shown in Figure 5.21. The significant effect in the expected direction is for *observe*. In all experimental conditions, the difference between the two points in time is significantly larger than in the control group (B: $t = 2.48$, $p = 0.017$, $d = 0.40$; C: $t = 2.45$, $p = 0.018$, $d = 0.38$; D: $t = 2.45$, $p = 0.018$, $d = 0.52$). Effect sizes are small for groups B and C, but the effect size is medium for group D. While on average the scale score does not change or

decreases somewhat in the experimental groups, it decreases more in the control group.

Statistically significant unexpected effects are found for scales *accept without judgement* and *emotional self-efficacy*. While in the control group the scale scores increase in T2 from T1, they decrease for group C for *accept without judgement* ($t=-2.46$, $p=0.018$, $d=-0.49$). Yet, the effect size is rather small. The scale score for *emotional self-efficacy* also increases in the control group increase, while the scale scores for groups B and D decrease. The differences are statistically significant and effect sizes are medium (B: $t=-2.44$, $p=0.020$, $d=-0.63$; D: $t=-2.44$, $p=0.020$, $d=-0.56$). Yet, this effect is inconsistent insofar as we would expect similar effects in both experimental groups where the same school staff programmes had been implemented and rather no effect in the experimental group with the student training.

5.3.22. Sweden: School staff's Relationship Skills and Social Awareness

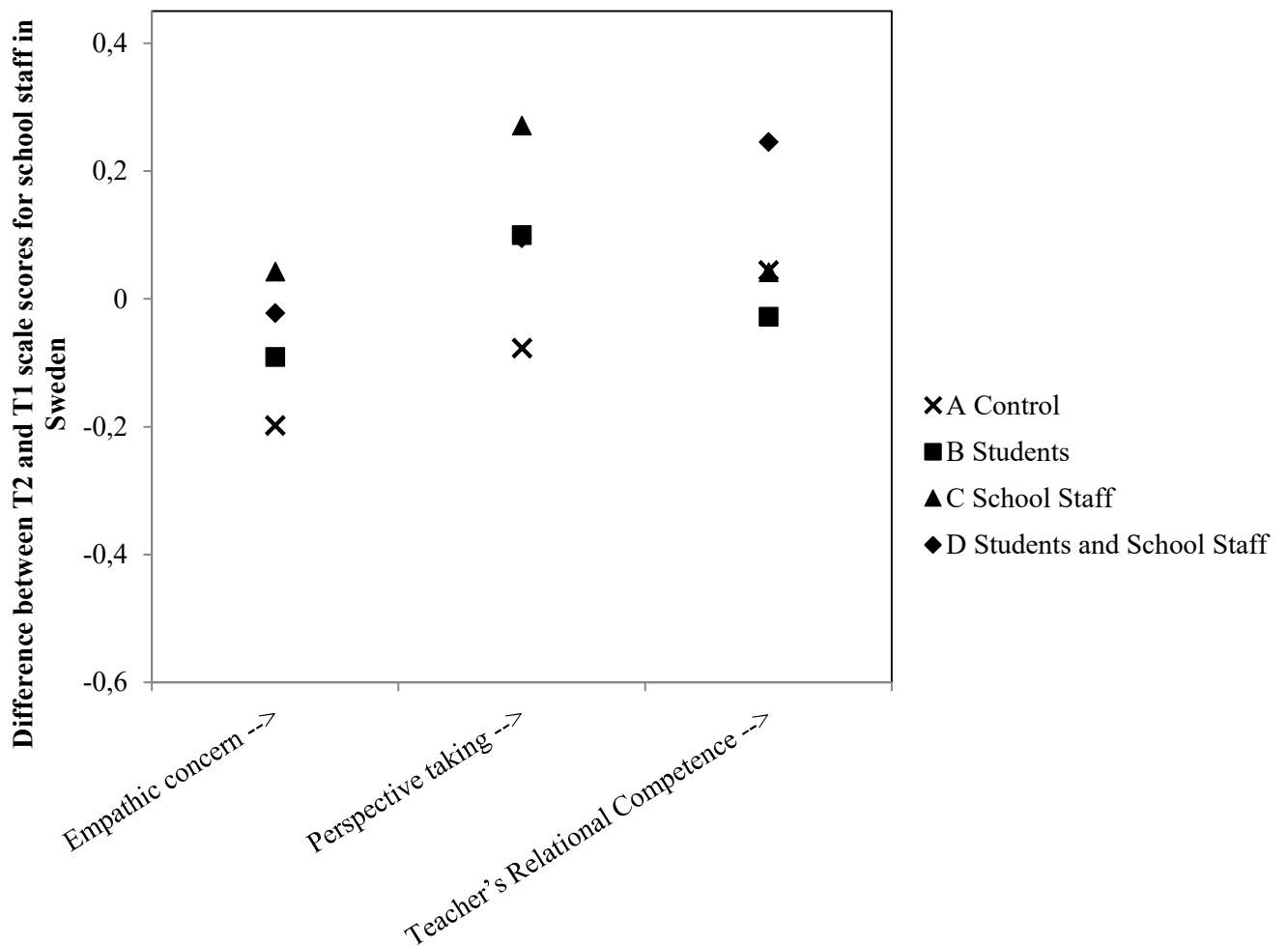


Figure 5.22. Average difference between T2 and T1 per group for the scales assessing the school staff's relationship skills and social awareness in Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

We find no significant effects for the relationship skills and social awareness scales for teachers in Sweden. Figure 5.22 presents the average differences between the two points in time for the control group and the three experimental groups and scales.

5.3.23. *Sweden: School staff's Intercultural Competencies/Diversity Awareness*

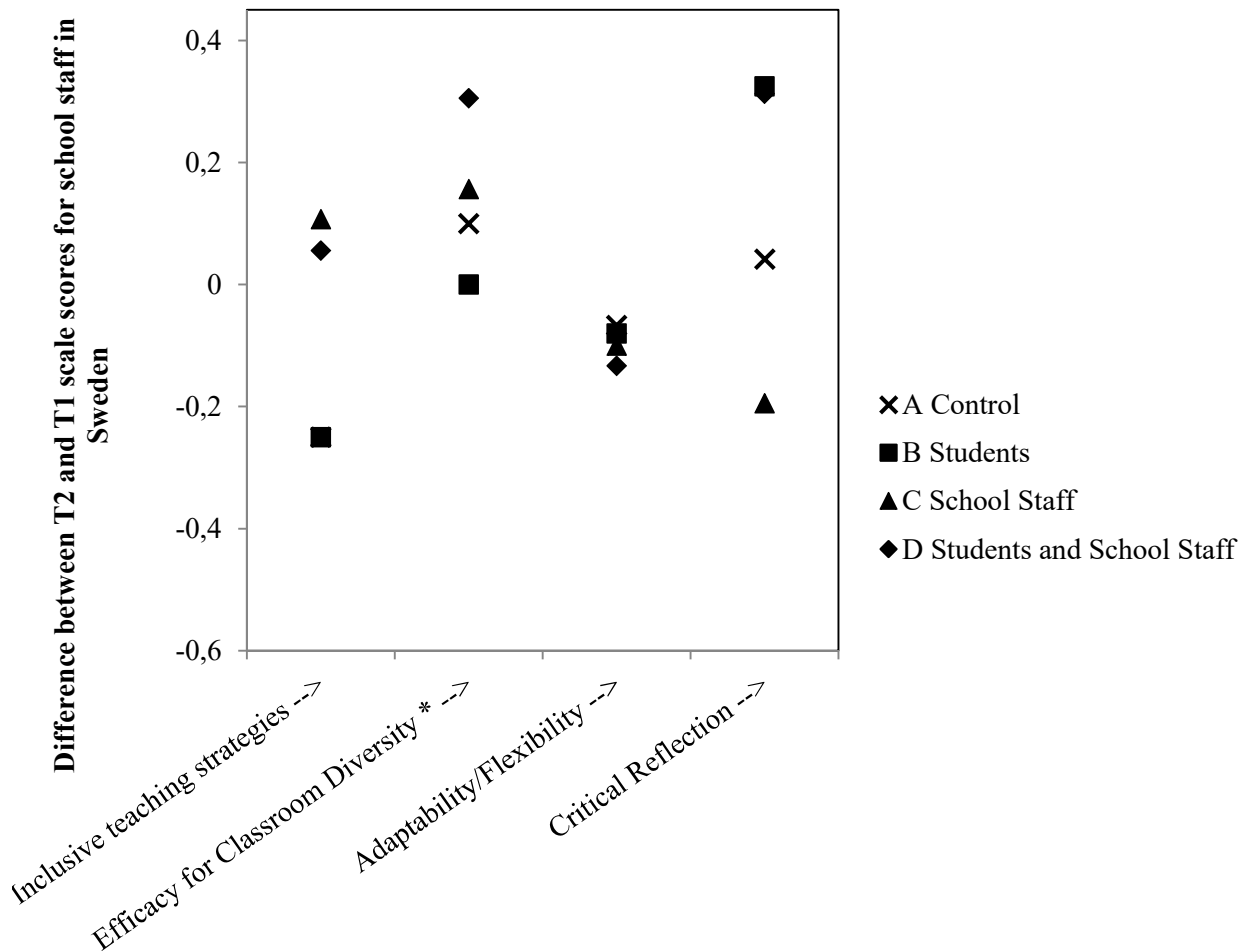


Figure 5.23. Average difference between T2 and T1 per group for the scales assessing the school staff's intercultural competencies/ diversity awareness in Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was between .60 and .70 in T1 and T2.

Although the mean scale values differ for some groups and certain scales that measure intercultural competencies/ diversity awareness (see Figure 5.23), none of the differences between the experimental and control groups is significant.

5.3.24. Summary of short-term effects

Table 5.1. presents an overview over short term-effects of the HAND in HAND programmes on students' self-reported social and emotional competencies and intercultural competencies/diversity awareness as well as on their cooperativeness reported by classmates (other report). Table 5.2. presents an overview over short-term effects of the HAND in HAND programmes on school staffs' self-reported social and emotional competencies and intercultural competencies/diversity awareness as well as on teachers' relational competence as perceived by students (other report). Altogether, the programmes had no short-term effects that were consistent across relevant experimental groups and school systems.

Looking at the results from the student scales we see more expected than unexpected effects in Slovenia, more unexpected than expected effects in Croatia in Sweden. More specifically, Table 5.1. suggests that there are only two outcomes that changed in the expected direction in both experimental groups where the HAND in HAND student programme had been implemented (experimental groups B and D): students' *self-control* in Sweden and students' *attitudes towards immigrants* in Slovenia. From this we can conclude that the student programme improved students' *self-control* in Sweden and *students' attitudes towards immigrants* in Slovenia. However, as we argued above, there is indication in the literature on social learning that whole-school approaches might be more effective, which would be supported by stronger effects in experimental group D than in experimental groups B and C. In accordance with this expectation we find an effect of a combination of all programmes on students' ability to *observe* in Croatia and for *empathic concern* und *perspective taking* in Slovenia. We also observe several negative effects of a combination of all student programmes: In Croatia the change of a *positive identity* is less positive as compared to the control group and there is a larger increase of *emotional problems* in group D. In Slovenia the ability to *accept without judgement* develops less positively in experimental group D than in the control group and in Sweden the *attitude towards immigrants*. However, none of these effects is consistent across school system and the effect sizes are small for all expected and unexpected effects.

Table 5.1.

Overview over short term-effects of the HAND in HAND programmes on students' self-reported social and emotional competencies and intercultural competencies/diversity awareness and as well as their cooperativeness reported by classmates (other report) in different experimental groups.

School System:	Croatia			Slovenia			Sweden		
Experimental group:	C	B	D	C	B	D	C	B	D
Self awareness									
Positive identity	no	no	-	no	-	no	+	no	no
Observe	+	no	+	no	+	no	no	-	no
Describe	no	no	no	no	no	no	no	no	no
Accept without judgement	no	no	no	no	no	-	no	no	no
Act with awareness	no	no	no	-	no	no	no	no	no
Self management									
Self control	no	no	-	no	-	no	+	+	+
Emotional problems	no	no	+	no	no	no	+	no	no
Aggressiveness	no	no	+	no	no	+	no	no	no
Relationship skills and social awareness									
Empathic concern	no	no	no	+	no	+	no	no	no
Perspective taking	no	no	no	+	no	no	no	no	no
Caring	no	no	no	no	no	+	no	no	no
Cooperation of peers (other report)	no	no	no	no	no	no	no	no	no
Intercultural competence/diversity awareness									
Attitudes towards immigrants	no	no	no	no	+	+	-	no	-
Critical Consciousness	no	no	no	no	no	no	no	no	no

Note: ++ means that the scale score for the respective outcome increased more or decreased less for the respective experimental group as compared to the control group and that the effect size was medium; + means that the scale score for the respective outcome increased more or decreased less for the respective experimental group as compared to the control and that the effect size was small; - means that the scale score for the respective outcome decreased more or increased less for the respective experimental group as compared to the control and that the effect size was small; -- means that the scale score for the respective outcome decreased more or increased less for the respective experimental group as compared to the control and that the effect size was medium; + or – printed in grey means that the effect size was very small; green means that the effect was in the expected direction, light green means that the effect was not necessarily expected but that it is plausible, red means that the effect was in the unexpected direction, light red means that the effect is rather implausible.

Table 5.2.

Overview over short term-effects of the HAND in HAND programmes on school staffs' self-reported social and emotional competencies and intercultural competencies/diversity awareness as well as teachers' relational competence as perceived by students (other report).

School System:	Croatia			Slovenia			Sweden		
Experimental group:	B	C	D	B	C	D	B	C	D
Self-awareness									
Observe	no	no	+	+	no	no	+	+	++
Describe	no	no	no	no	no	no	no	no	no
Accept without judgement	no	no	no	no	no	no	no	-	no
Act with awareness	no	no	no	-	no	-	no	no	no
Self-management									
Emotional self-efficacy	no	no	no	no	no	no	--	no	--
Emotional problems	no	no	no	no	-	no	no	no	no
Relationship skills and social awareness									
Empathic concern	no	no	no	+	no	+	no	no	no
Perspective taking	no	no	no	no	no	no	no	no	no
Teachers' relational competence (self-report)	no	no	no	no	-	no	no	no	no
Teachers' relational competence (student report)	-	no	no	no	no	no	no	no	no
Intercultural competence/diversity awareness									
Inclusive teaching strategies (self-report)	no	no	no	no	no	no	no	no	no
Self-efficacy for addressing classroom diversity	no	no	no	no	no	+	no	no	no
Adaptability/Flexibility	no	no	no	no	no	no	no	no	no
Critical Reflection	no	-	no	no	no	no	no	no	no

Note: ++ means that the scale score for the respective outcome increased more or decreased less for the respective experimental group as compared to the control group and that the effect size was medium; + means that the scale score for the respective outcome increased more or decreased less for the respective experimental group as compared to the control and that the effect size was small; - means that the scale score for the respective outcome decreased more or increased less for the respective experimental group as compared to the control and that the effect size was small; -- means that the scale score for the respective outcome decreased more or increased less for the respective experimental group as compared to the control and that the effect size was medium; + or – printed in grey means that the effect size was very small; green means that the effect was in the expected direction, light green means that the effect was not necessarily expected but that it is plausible, red means that the effect was in the unexpected direction, light red means that the effect is rather implausible.

With regard to effects on school staffs' social and emotional competencies and intercultural competencies/diversity awareness we find more expected than unexpected effects in Slovenia and equal numbers of expected and unexpected effects in Croatia and Sweden (see Table 5.2). There is only one outcome that changed in the expected direction in both experimental groups that had participated in the HAND in HAND programmes for school staff: *observe* in Sweden. In Croatia also an effect on *observe* is observed, but only for group D when student and staff programmes were combined. In Slovenia no effect of the HAND in HAND programme on the *observe*-scale was found. Here significant, yet small, effects on *empathic concern* and on *self-efficacy for addressing diversity* are found which both only exist in group D and group B but not in group C. Hence, in Croatia the HAND in HAND programme appears to have an effect on school staffs' ability to observe – but only when the programmes addressing students and school staff are all implemented in the school. In Sweden the HAND in HAND programme also appears to have an effect on school staffs' ability to observe – no matter whether the programmes are combined or whether only of staff programmes were implemented in the school. Additionally, a negative effect of the student programme on teachers' *emotional self-efficacy* is observed. In Slovenia a combination of student and school staff programmes might have a positive short-term effect on school staffs' *empathic concern* and on *self-efficacy for addressing diversity*, but a negative short-term effect on *act with awareness*. Yet, it should be noted, that none of these effects were consistent across school systems and that most of the effect sizes were small.

It can be concluded that several positive and negative effects of the HAND in HAND programmes are suggested by our analyses. However, most effect sizes are small and the effects differ considerably between school systems.

Table 5.2.

Overview over short term-effects of the HAND in HAND programmes on school staffs' self-reported social and emotional competencies and intercultural competencies/diversity awareness as well as teachers' relational competence as perceived by students (other report).

School System:	Croatia			Slovenia			Sweden		
Experimental group:	B	C	D	B	C	D	B	C	D
Self-awareness									
Observe	no	no	+	+	no	no	+	+	++
Describe	no	no	no	no	no	no	no	no	no
Accept without judgement	no	no	no	no	no	no	no	-	no
Act with awareness	no	no	no	-	no	-	no	no	no
Self-management									
Emotional self-efficacy	no	no	no	no	no	no	--	no	--
Emotional problems	no	no	no	no	-	no	no	no	no
Relationship skills and social awareness									
Empathic concern	no	no	no	+	no	+	no	no	no
Perspective taking	no	no	no	no	no	no	no	no	no
Teachers' relational competence (self-report)	no	no	no	no	-	no	no	no	no
Teachers' relational competence (student report)	-	no	no	no	no	no	no	no	no
Intercultural competence/diversity awareness									
Inclusive teaching strategies (self-report)	no	no	no	no	no	no	no	no	no
Self-efficacy for addressing classroom diversity	no	no	no	no	no	+	no	no	no
Adaptability/Flexibility	no	no	no	no	no	no	no	no	no
Critical Reflection	no	-	no	no	no	no	no	no	no

Note: ++ means that change in the respective outcome was stronger positive or weaker negative in the respective experimental group as compared to the control group and that the effect size was medium; + means that change in the respective outcome was stronger positive or weaker negative in the respective experimental group as compared to the control group and that the effect size was small; - means that change in the respective outcome was stronger negative or weaker positive in the respective experimental group as compared to the control group and that the effect size was small; -- means that change in the respective outcome was stronger negative or weaker positive in the respective experimental group as compared to the control group and that the effect size was medium; + or – printed in grey means that the effect size was very small; green means that the effect was in the expected direction, light green means that the effect was not necessarily expected but is plausible, red means

that the effect was in the unexpected direction, light red means that the effect was not necessarily unexpected but is rather implausible.

5.4. Middle-term programme effects

In the following, programme effects on changes in students' and teachers' self-reported social and emotional competencies as well as intercultural competencies/diversity awareness between T1 and T3 are presented.

5.4.1. Croatia: Students' Self-awareness

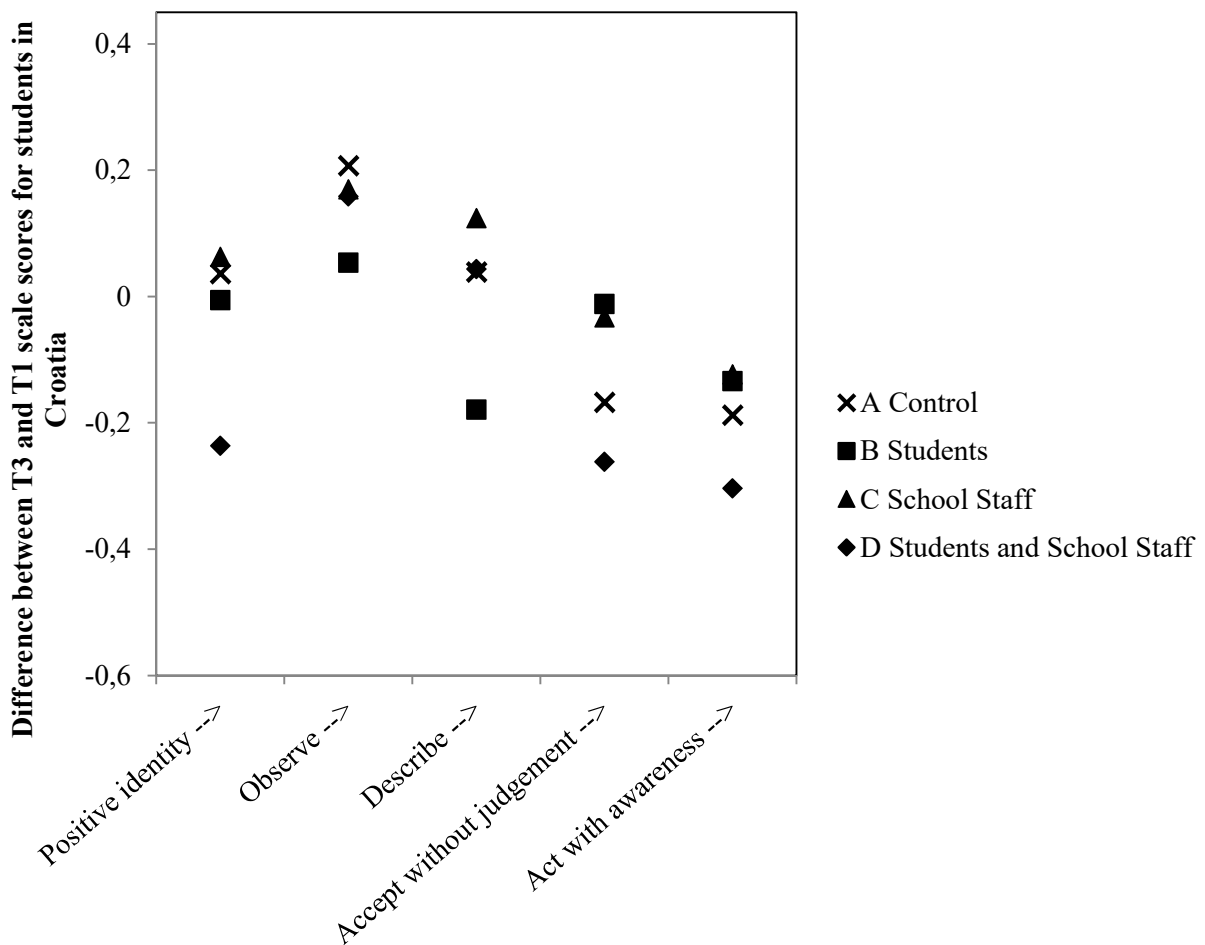


Figure 5.24. Average difference between T3 and T1 per group for the scales assessing the students' self-awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average manifest change for the three experimental groups and the control group for scales measuring aspects of self-awareness are shown in Figure 5.24. Only one of the 15 differences between the control group and experimental groups was significant. The effect is observed in *positive identity* and points into the

unexpected direction. The difference in the change score between the control group and condition D is significant ($t = -4.04$, $p = 0.000$, $d = -0.40$). For students from group D a lower level of positive identity at T3 compared to T1 was observed while the level in the control group only changed slightly between these two points in time. However, the effect size is small. Moreover, experimental group B that also participated in the student programme just like group D, showed no change.

5.4.2. Croatia: Students' Self-management

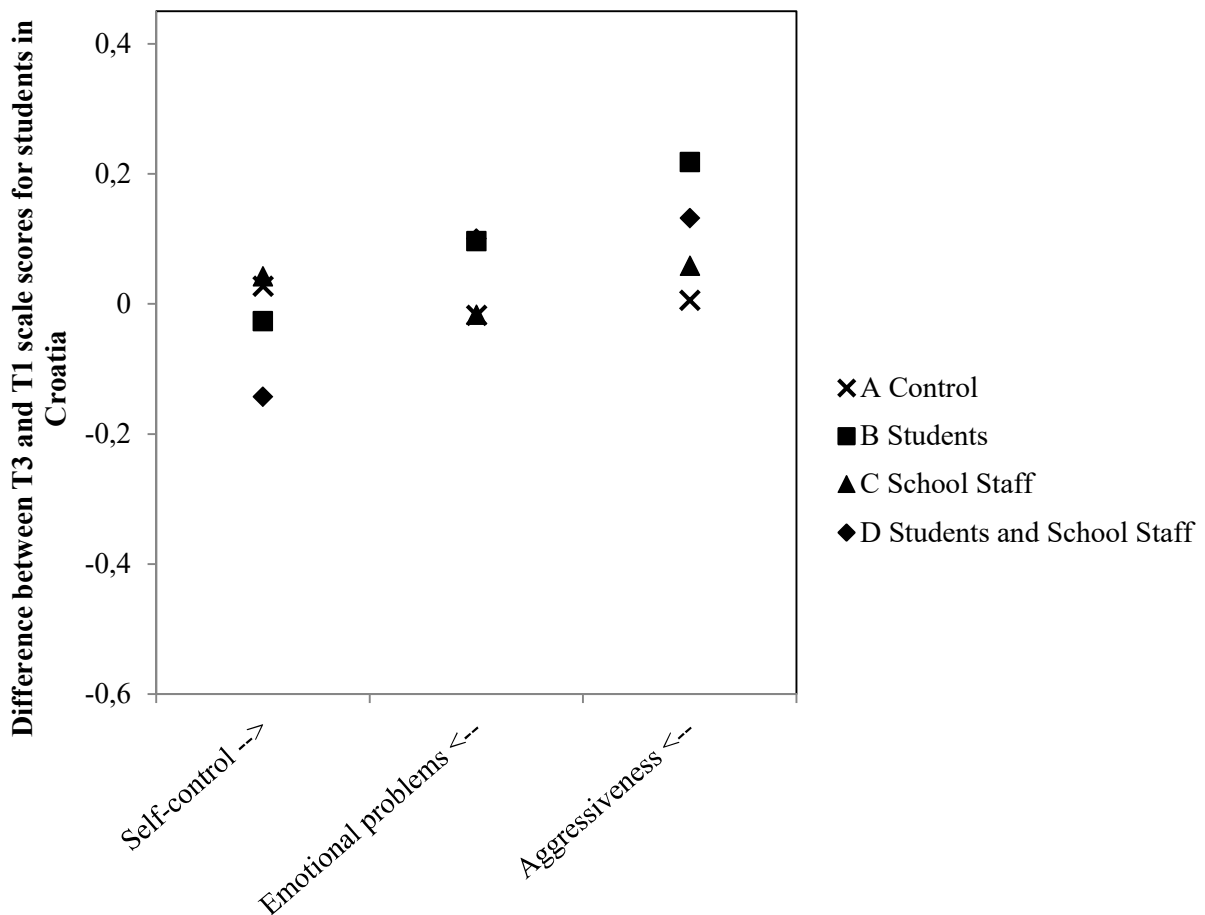


Figure 5.25. Average difference between T3 and T1 per group for the scales assessing the students' self-management in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

We find three significant effects in the unexpected direction for the self-management scales in Croatia. The average differences for the control group and

the three experimental groups per scale between T3 and T1 are shown in Figure 5.25. For the *self-control* scale, group D significantly differs from the control group ($t=-2.59$, $p=0.010$, $d=-0.29$). For students in the condition where students and school staff were exposed to the programme we observe on average a larger decrease in scale scores compared to the control group. However, the effect size is small and a similar effect was not found for experimental group B that had received the same student programme. In *emotional problems* and *aggressiveness*, group B significantly differs from the control group, but the effect size is again small (*emotional problems*: $t=2.26$, $p=0.025$, $d=0.16$; *aggressiveness*: $t=3.76$, $p=0.000$, $d=0.32$). For students subject to the condition where only students were exposed to the programme we observe on average a larger increase in scale scores compared to the control group. For group D there is also an increase in emotional problems and aggressiveness, but the difference to the control group is not statistically significant.

5.4.3. Croatia: Students' Relationship Skills and Social Awareness

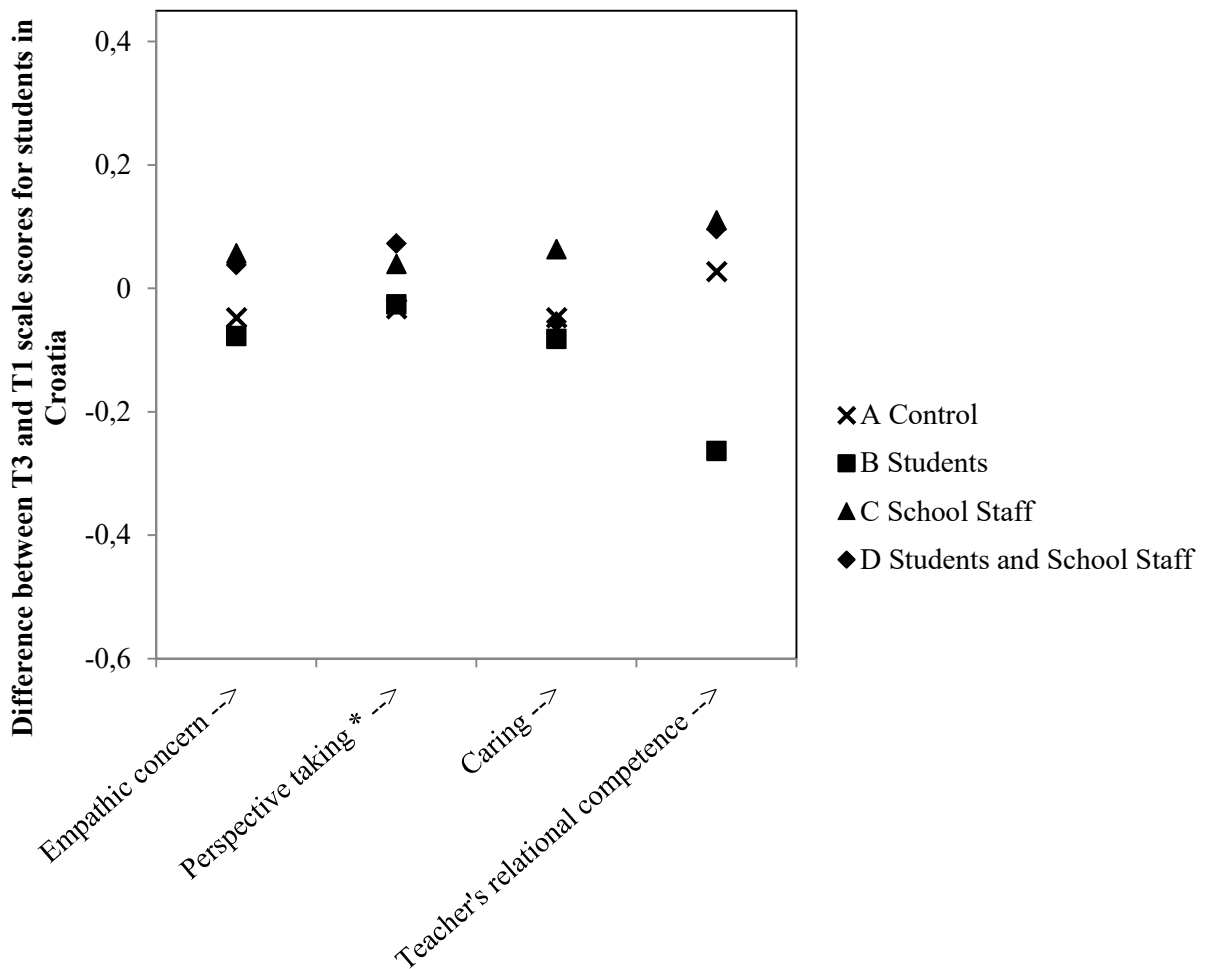


Figure 5.26. Average difference between T3 and T1 per group for the scales assessing the students' relationship skills and social awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was between .60 and .70 in T1.

There is one significant difference between the experimental groups and the control group for the scales that were included to measure relationship skills and social awareness in Croatia. The average differences between the two points in time for the control group and the three experimental groups and scales are presented in Figure 5.26. For *teacher's relational competence* the students from condition D reported a significantly different manifest change as compared to the control group ($t=2.03$, $p=0.043$, $d=0.09$). The effect points into the expected direction, but the effect size is very small. The scale scores in condition D increase while the scale scores for the control group do not change much. No such effect

is observed for group B that participated in a similar training as group D. Here, even a large decrease is visible, but the difference to the control group is not significant.

5.4.4. Croatia: Students' Cooperation (Other-reports)

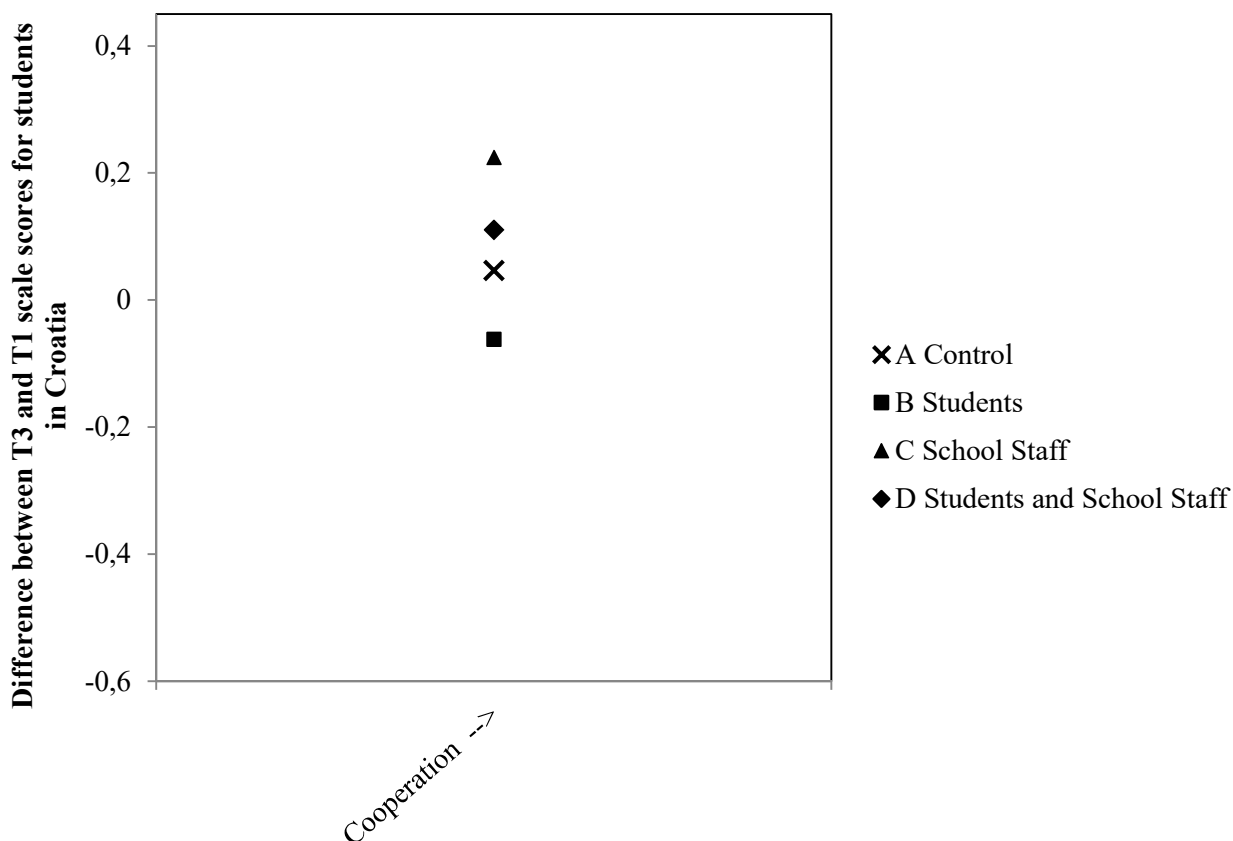


Figure 5.27 Average difference for different conditions between T3 and T1 in students' scale score for cooperation in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences for the conditions for cooperative abilities are presented in Figure 5.27. We find no significant effect for *cooperation* in Croatia.

5.4.5. Croatia: Students' Intercultural Competencies/Diversity Awareness

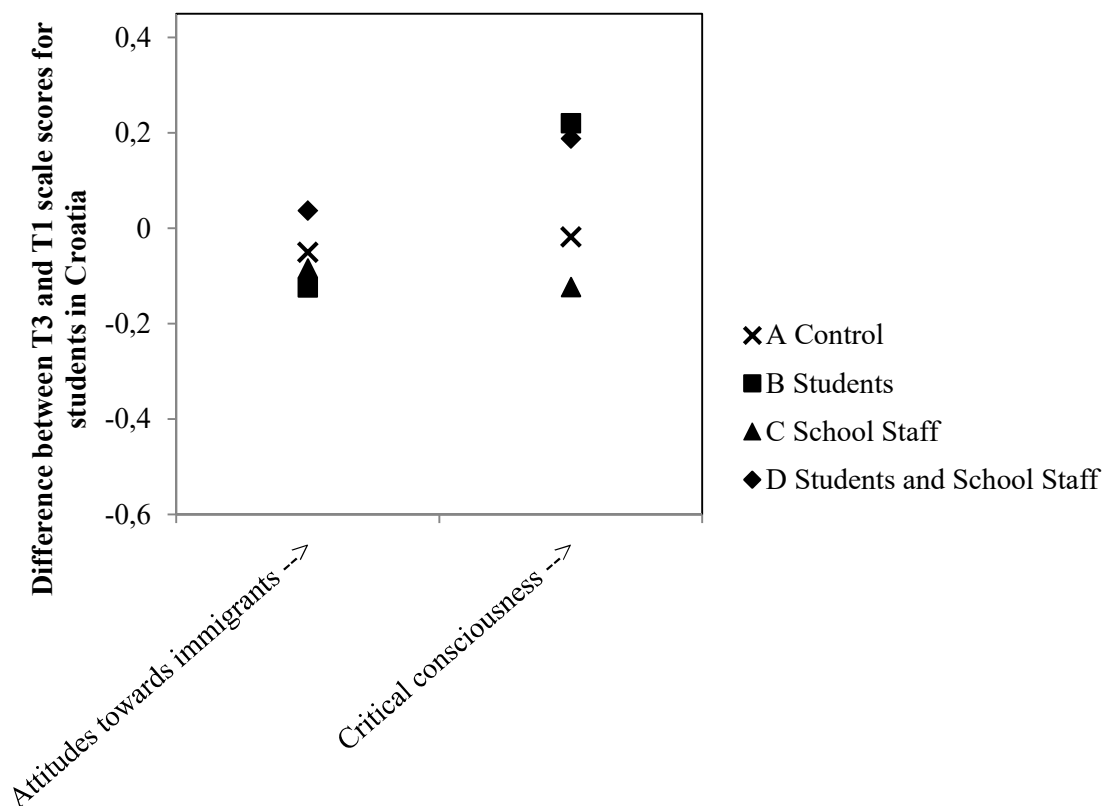


Figure 5.28. Average difference between T3 and T1 per group for the scales assessing the students' intercultural competencies/ diversity awareness in Croatia
Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure intercultural competencies/ diversity awareness are shown in Figure 5.28. No significant differences between the control and experimental groups are found for these scales.

5.4.6. Croatia: School Staff's Self-awareness and Self-management

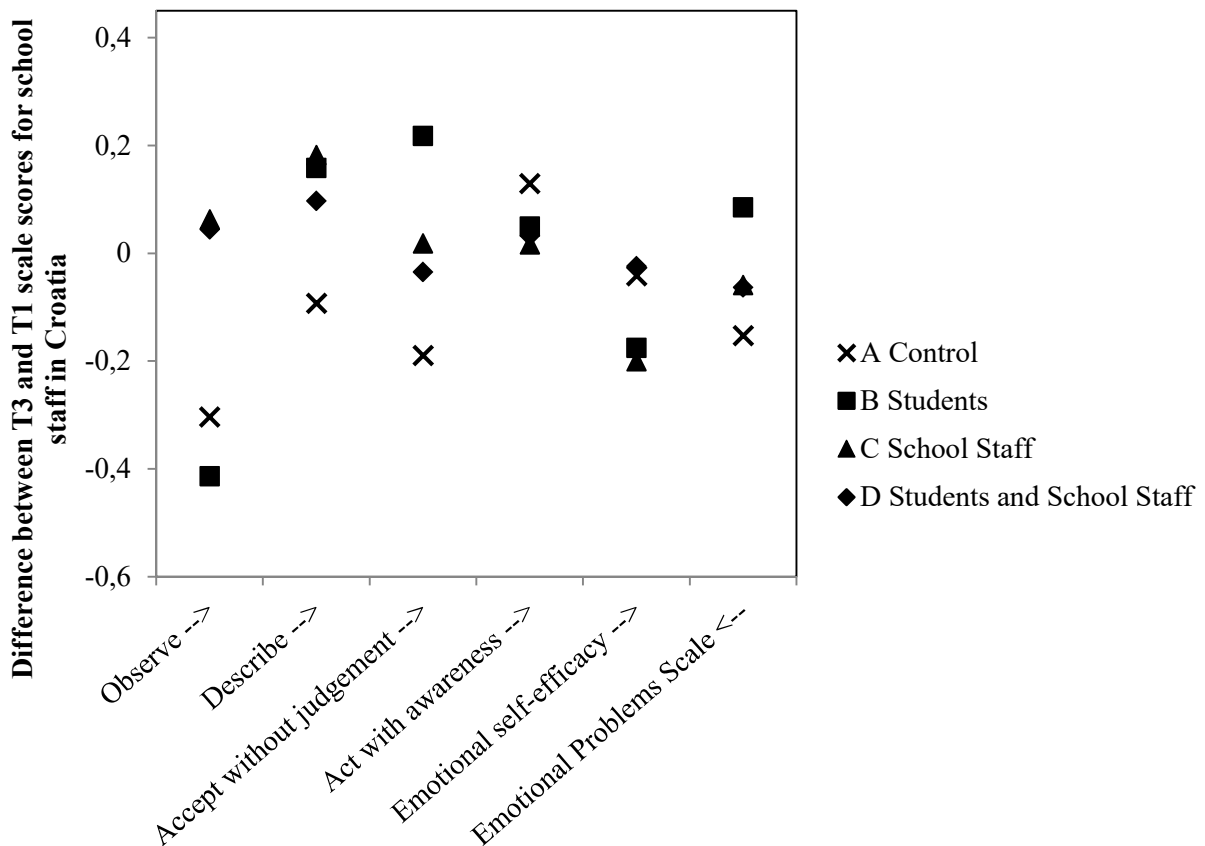


Figure 5.29. Average difference between T3 and T1 per group for the scales assessing self-awareness and self-management for school staff in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figure 5.29 presents the average differences between the two points in time for the control group and the three experimental groups and scales for self-awareness and self-management. We find significant effects in four scales. An effect in the expected direction is found for *observe* in condition C. The difference between T3 and T1 is, on average, larger in condition C than in the control group ($t=3.17$, $p=0.002$, $d=0.47$). We observe a decrease between T3 and T1 in the control group, while the average scale score in group C does not change. However, the effect size is small and no such difference to the control group is observed for condition D. Also, we observe an effect in the expected direction for the scale *describe*. The changes in groups C and D are significantly larger than in the control group – yet the effect sizes are small (C: $t=3.22$, $p=0.002$, $d=0.40$: D: $t=3.24$, $p=0.002$, $d=0.28$). In addition, we observe one plausible but not necessarily expected effect for *accept without judgement*. The manifest positive change is significantly larger

for group B in comparison to the control group and the effect sizes is medium ($t=3.21, p=0.002, d=0.51$). Yet, no school staff training had taken place in group B.

Regarding *emotional self-efficacy* we observe an effect in the unexpected direction. In group C the change is significantly different from the one in the control group ($t=-2.46, p=0.017, d=-0.30$). Emotional self-efficacy declines, on average, from T1 to T3 significantly more in schools with only school staff training than in the control group schools. However the effect size is small and no such effect is observed in group D, i.e. in schools where school staff programmes and a student programme had been implemented.

5.4.7. Croatia: School Staff's Relationship Skills and Social Awareness

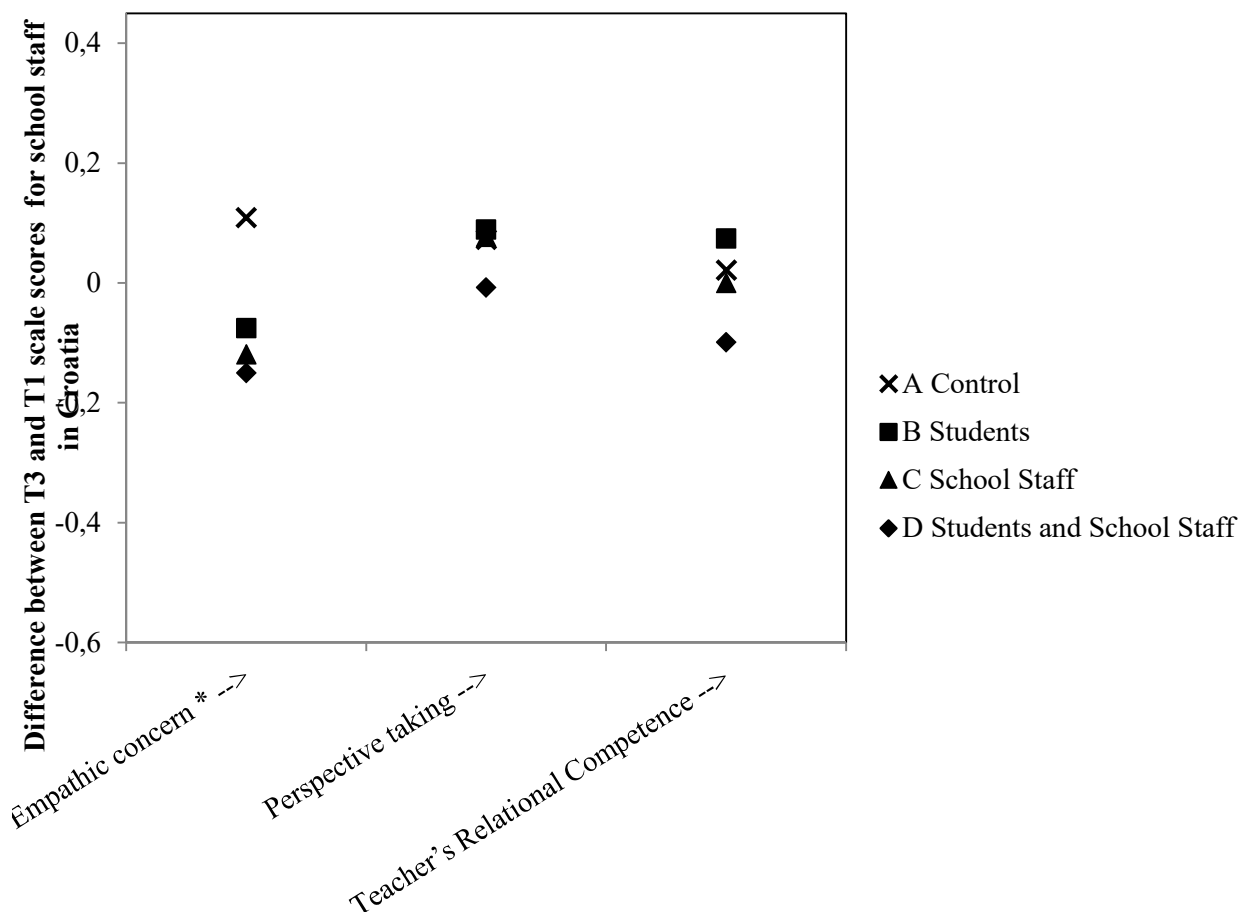


Figure 5.30. Average difference between T3 and T1 per group for the scales assessing the school staff's relationship skills and social awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

* Reliability of this scale was between .60 and .70 in T3.

We find one significant effect pointing in the unexpected direction for relationship skills and social awareness. The average differences between the two points in time for the control group and the three experimental groups and scales are presented in *Figure 5.30*. A significant unexpected effect was found for *empathic concern* in group D ($t=-2.28, p=0.026, d=-0.34$) in comparison to the control group: Empathic concern improves in the control condition but in group D (students and school staff programmes) it decreases on average. However, the effect size is small and no such difference was observed between the control group and group C (school staff programmes).

5.4.8. Croatia: School Staff's Intercultural Competencies/Diversity Awareness

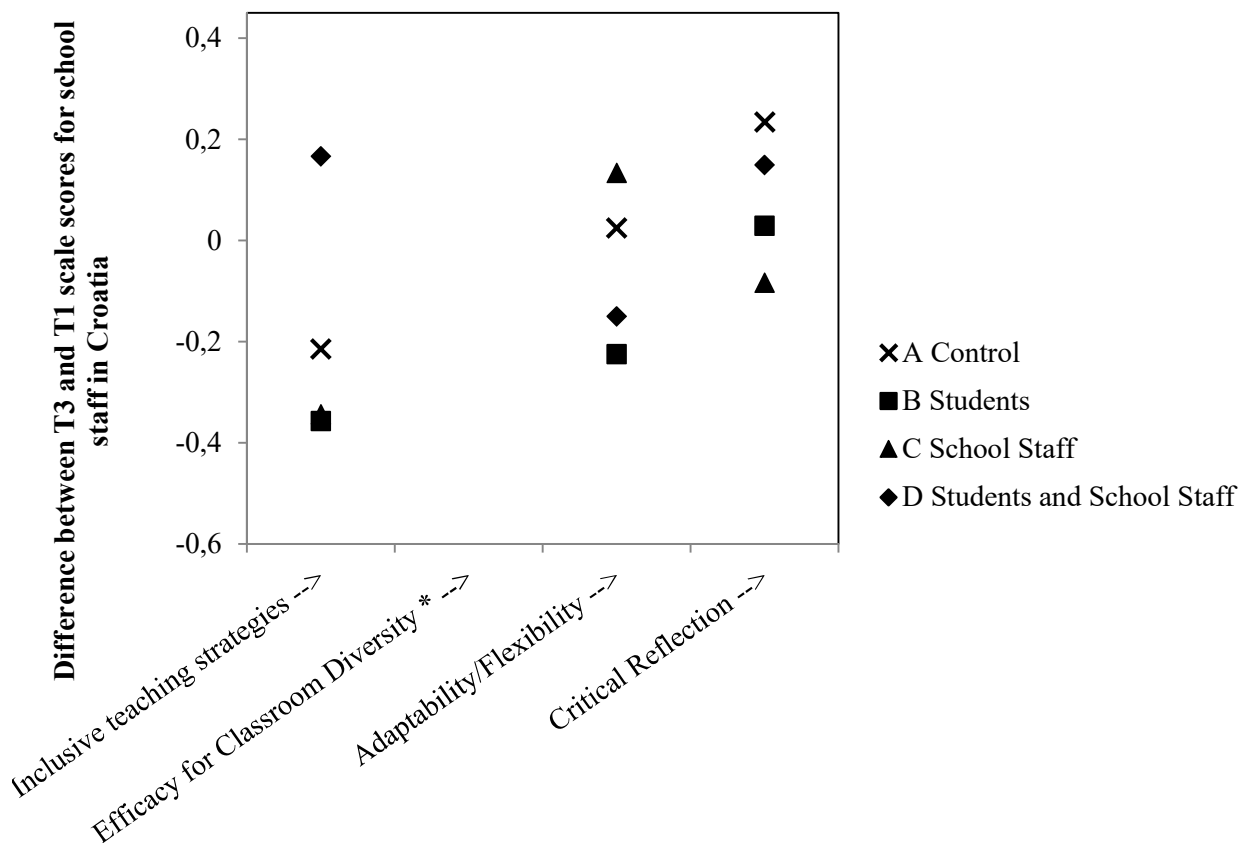


Figure 5.31. Average difference between T3 and T1 per group for the scales assessing the school staff's intercultural competencies/ diversity awareness in Croatia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was too low to be reported.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure intercultural competencies/diversity awareness are presented in *Figure 5.31*. One significant effect in the unexpected direction is found. The difference is found for *critical reflection*. In group C (but not in group D) a decrease of the scale scores was observed whereas in the control group the scale scores increased ($t=-3.37, p=0.001, d=-0.46$). The effect size is small.

5.4.9. Slovenia: Students' Self-awareness

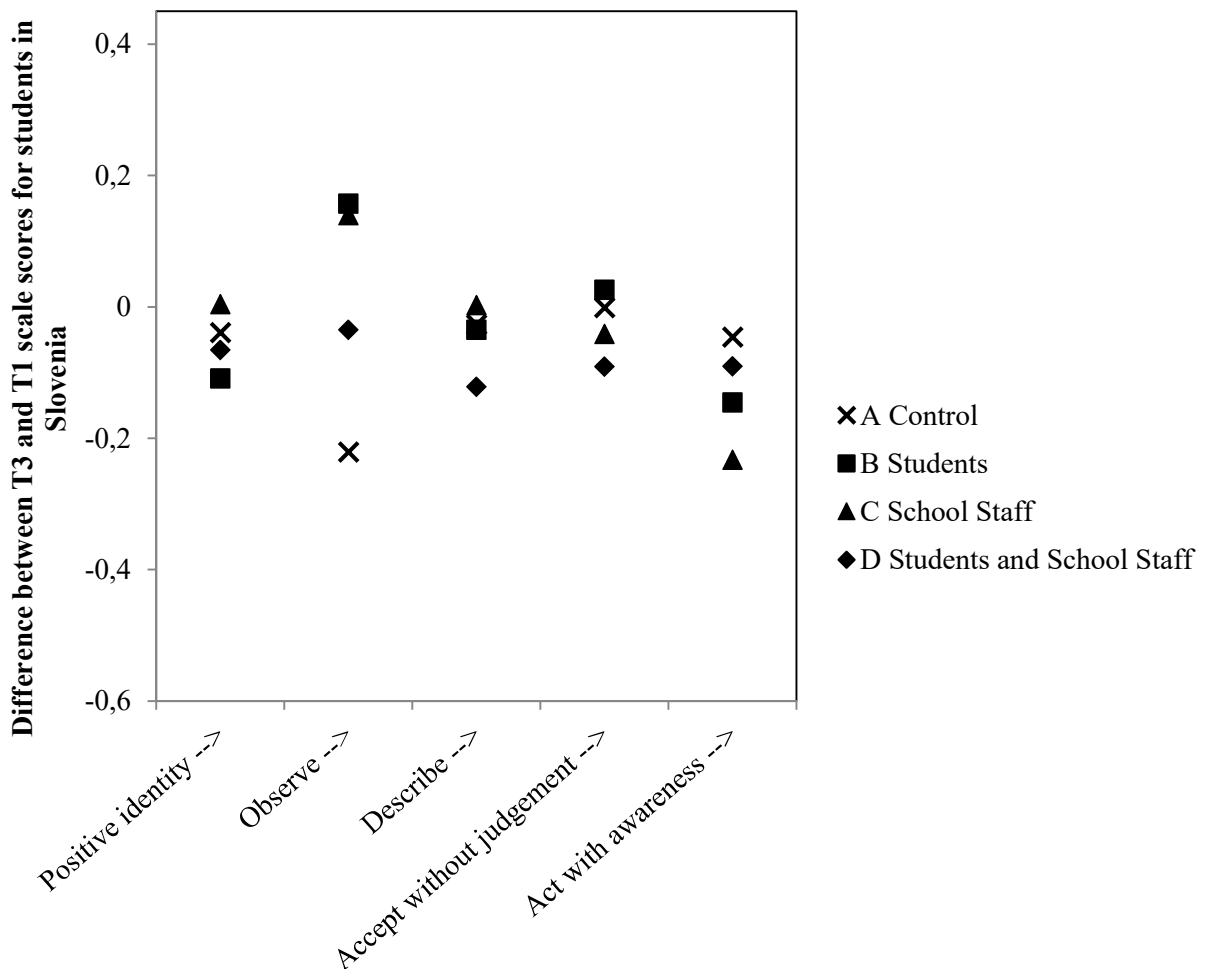


Figure 5.32. Average difference between T3 and T1 per group for the scales assessing the student's self-awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences for the control group and the three experimental groups and scales that measure self-awareness are shown in *Figure 5.32*. In Slovenia, we can observe one expected effect for the *observe* scale. Conditions B, C and D significantly differ from the control group (B: $t=4.29, p=0.000, d=0.43$; C: $t=5.42, p=0.000, d=0.41$; D: $t=2.44, p=0.015, d=0.22$). In the control group the scale score decreased between T1 and T3, it remained the same in Condition D and increased in Conditions B and C. However, the effect sizes are small.

Moreover, we find one effect in the unexpected direction for *act with awareness*. Group B (student programme) significantly differs from the control group ($t = -2.18, p=0.030, d=-0.13$). The scores in the experimental group B (but not in experimental group D with student programme plus staff programmes) decrease on average, whereas in the control group they only change slightly. Yet, the effect size is very small.

5.4.10. Slovenia: Students' Self-management

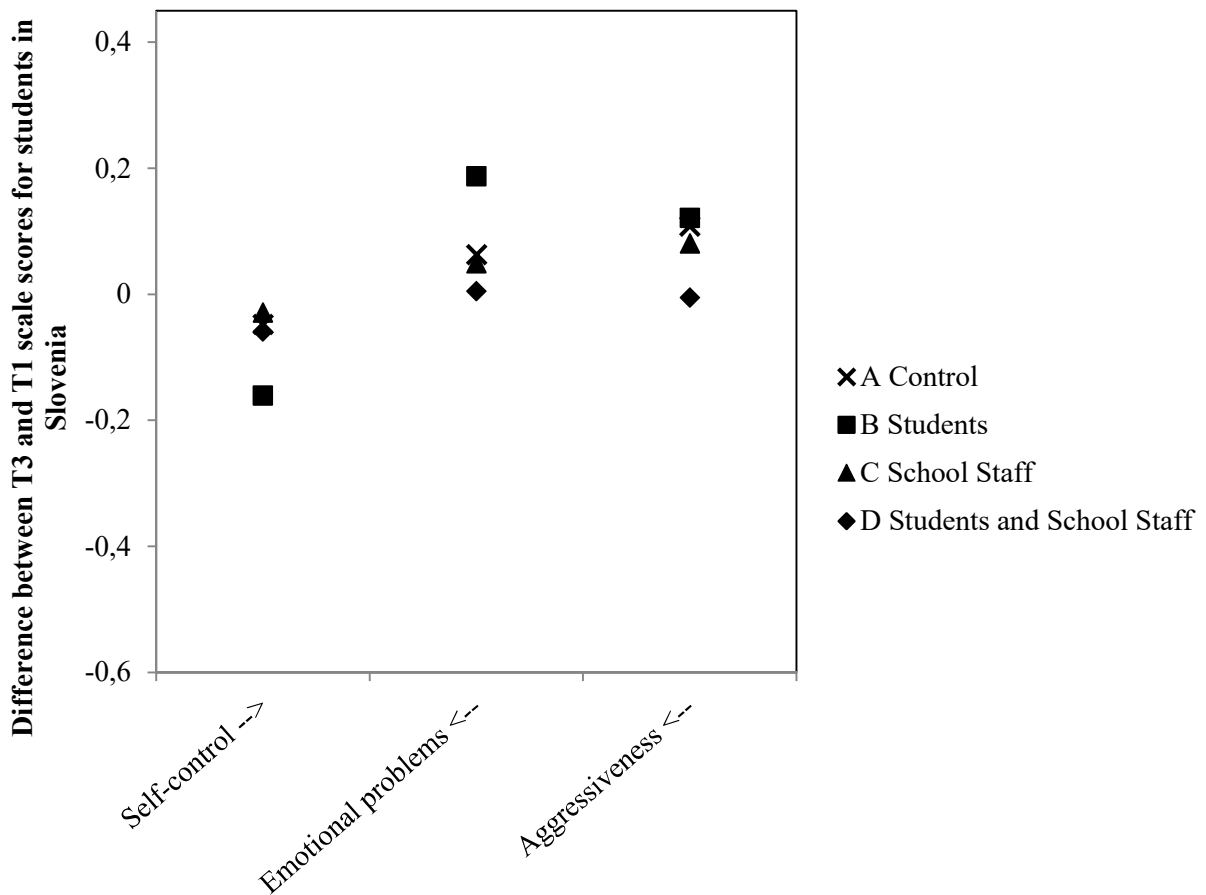


Figure 5.33. Average difference between T3 and T1 per group for the scales assessing the student's self-management in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figure 5.33 presents the average differences for the control group and the three experimental groups and scales targeting self-management in the student questionnaire in Slovenia. There is one significant effect for these scales in the unexpected direction. We observe a significantly larger decrease in the scale scores for students' self-reported *self-control* in group B (student programme) in comparison to the control group ($t=-3.62, p=0.000, d=-0.19$). However, the effect size is very small and no such effect is observed for group D (student programme and school staff programmes).

5.4.11. Slovenia: Students' Relationship Skills and Social Awareness

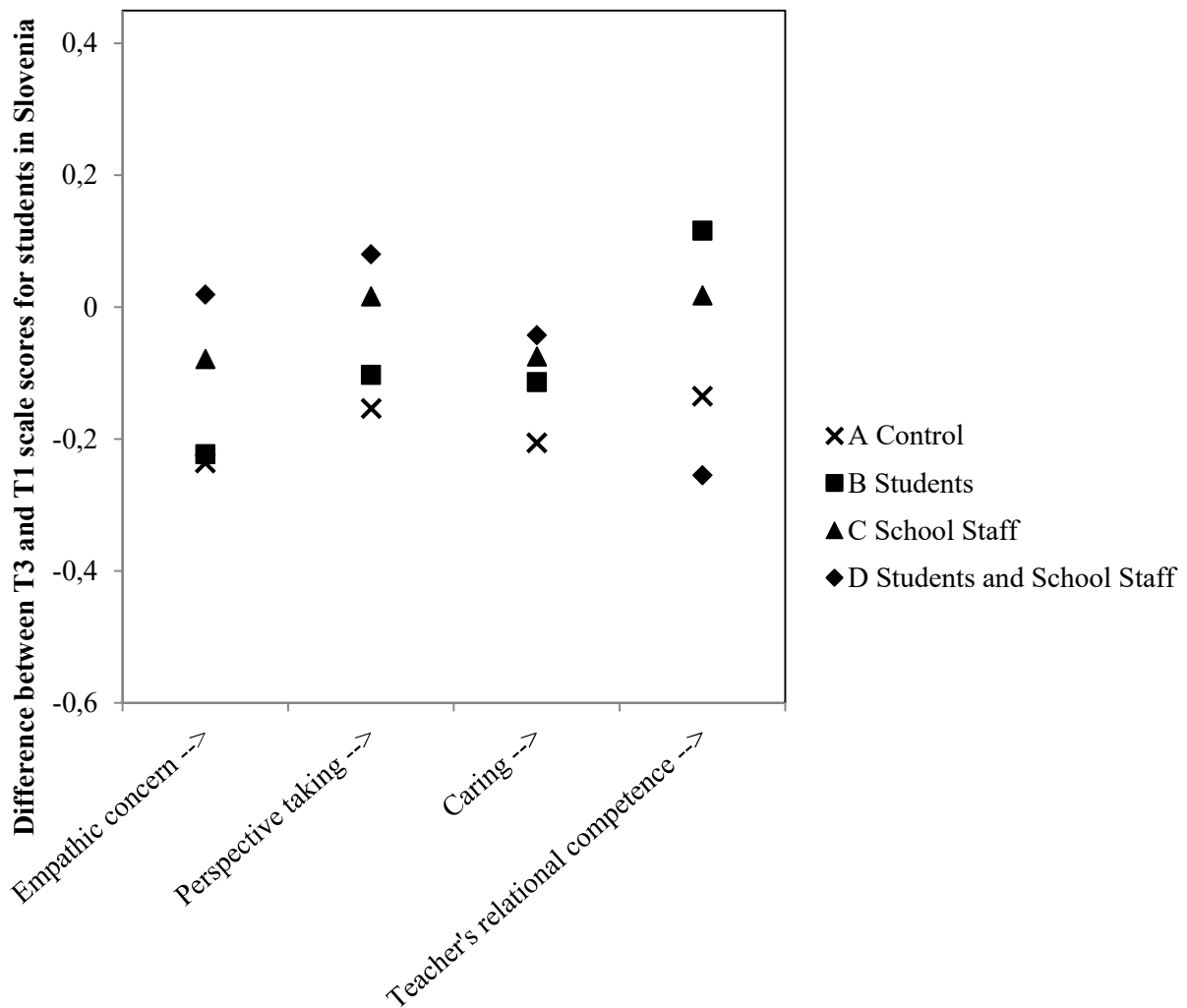


Figure 5.34. Average difference between T3 and T1 per group for the scales assessing the student's relationship skills and social awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figure 5.34 presents the average differences for the control group and the three experimental groups and scales that measure students' relationship skills and social awareness. We find effects in the expected direction for *empathic concern* and *perspective taking*. The increases for groups C and D are significantly larger than in the control group (*empathic concern*: C: $t=2.06$, $p=0.040$, $d=0.21$; D: $t=3.70$, $p=0.000$, $d=0.34$; *perspective taking*: C: $t=2.18$, $p=0.030$, $d=0.23$; D: $t=3.10$, $p=0.002$, $d=0.32$). However, the effect sizes are all small and no such effect is observed for group B where only students participated in the HAND in HAND student programme. For *teacher's relational competence* we find another

expected effect for condition B and an unexpected effect for condition D. While in the control group the scores decreased they decreased significantly more in condition D ($t=-2.02$, $p=0.045$, $d=-0.14$). However, the effect size is very small. The scores increased in condition B ($t=3.81$, $p=0.000$, $d=0.32$). The effect size is small.

5.4.12. Slovenia: Students' Cooperation

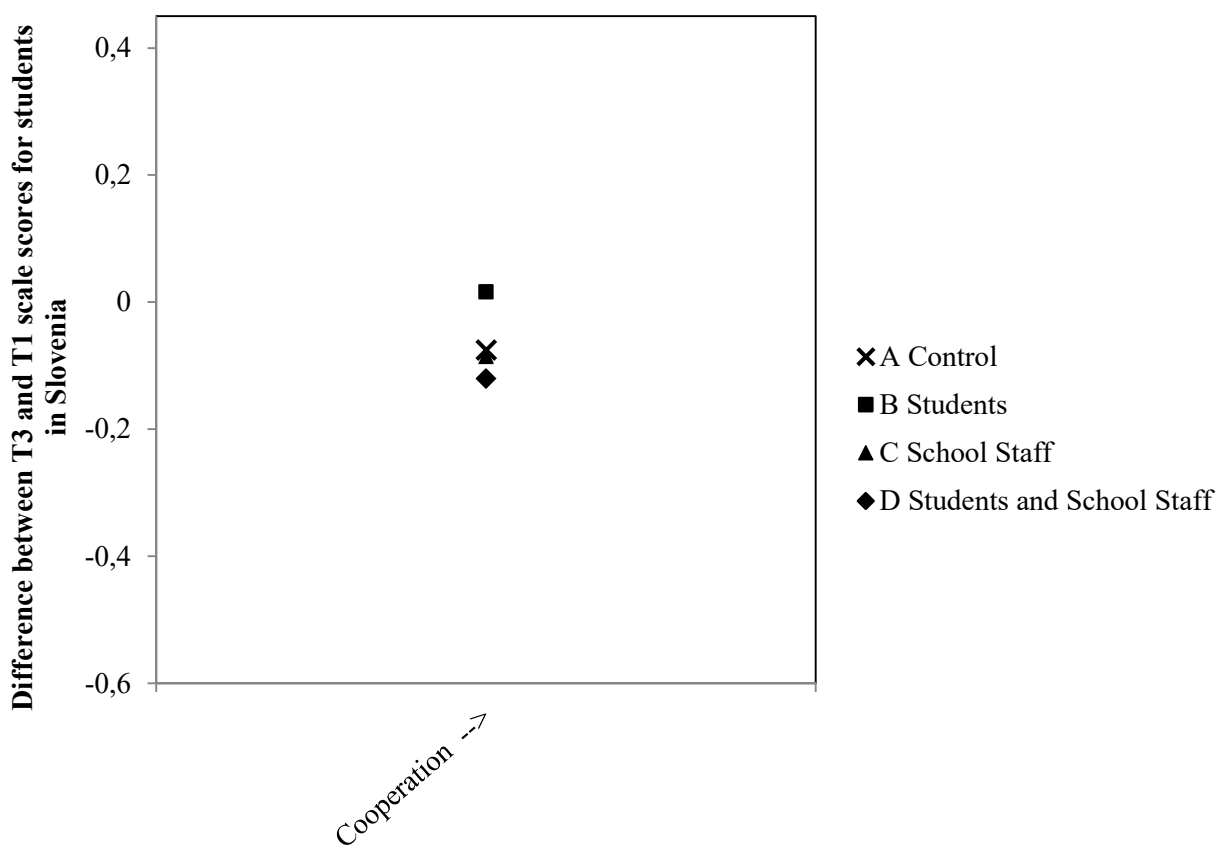


Figure 5.35 Average difference for different conditions between T3 and T1 in students' scale score for cooperation in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences for the conditions in *cooperation* are presented in Figure 5.35. We find no effect in *cooperation* for students in Slovenia.

5.4.13. Slovenia: Students' Intercultural Competencies/Diversity Awareness

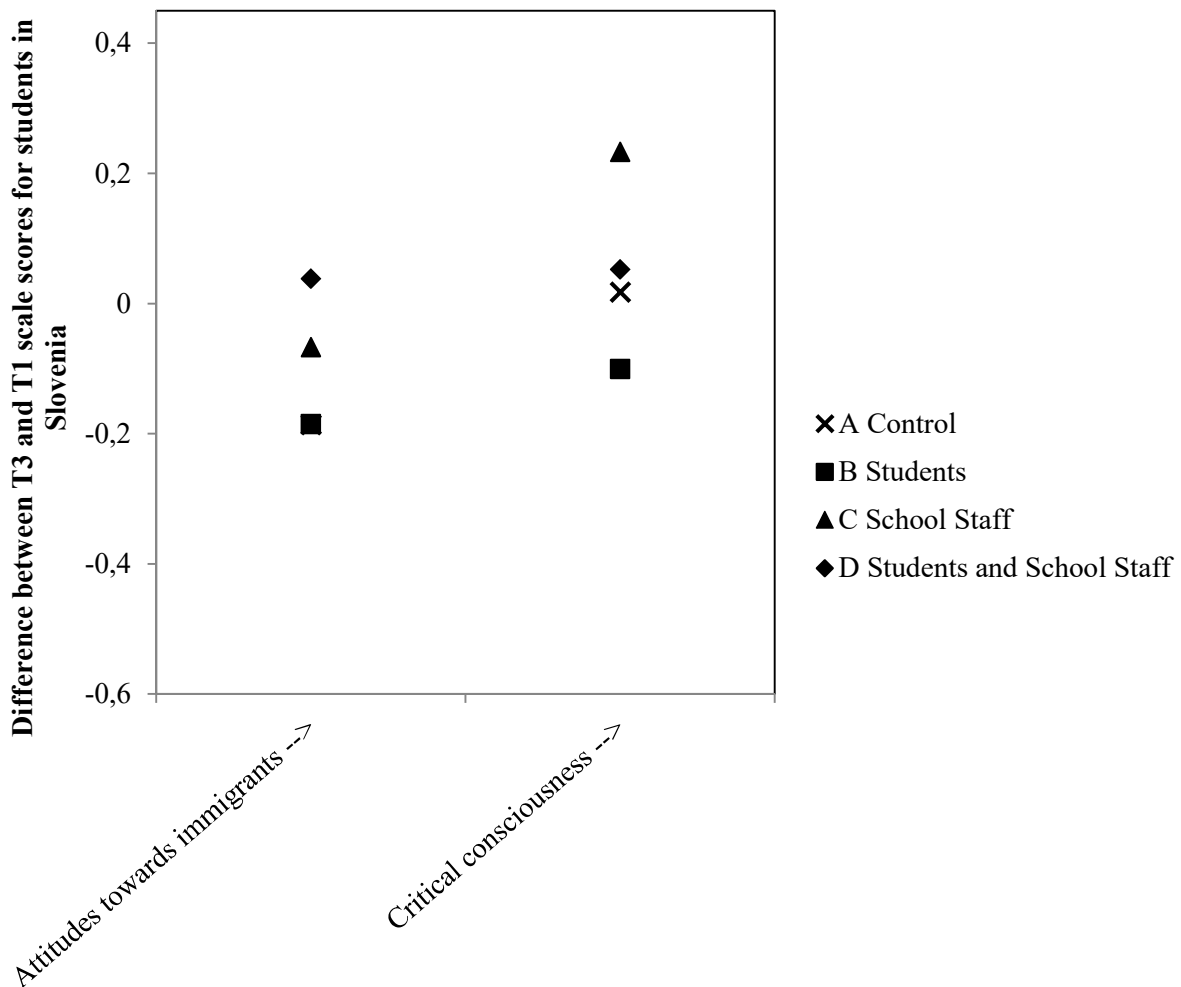


Figure 5.36. Average difference between T3 and T1 per group for the scales assessing the student's intercultural competencies/ diversity awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure students' intercultural competencies/ diversity awareness are presented in Figure 5.36. We establish a significant effect in the expected direction for students' self-reported *attitudes towards immigrants*. The difference between the two points in time is significantly larger in groups C and D than in the control group (C: $t=3.13$, $p=0.002$, $d=0.17$, D: $t=4.48$, $p=0.000$, $d=0.33$). While the level of positive attitudes towards immigrants falls from T1 to T3 in the control group, it changes only slightly in experimental groups C and D. Yet the effects are very small and small respectively. Another effect can be observed for *critical consciousness*. In

group C the scores increase while in the control group they stay constant ($t=3.01$, $p=0.003$, $d=0.25$). Again, the effect size is only small and group D does not differ from the control group.

5.4.14. Slovenia: School staff's Self-awareness and Self-management

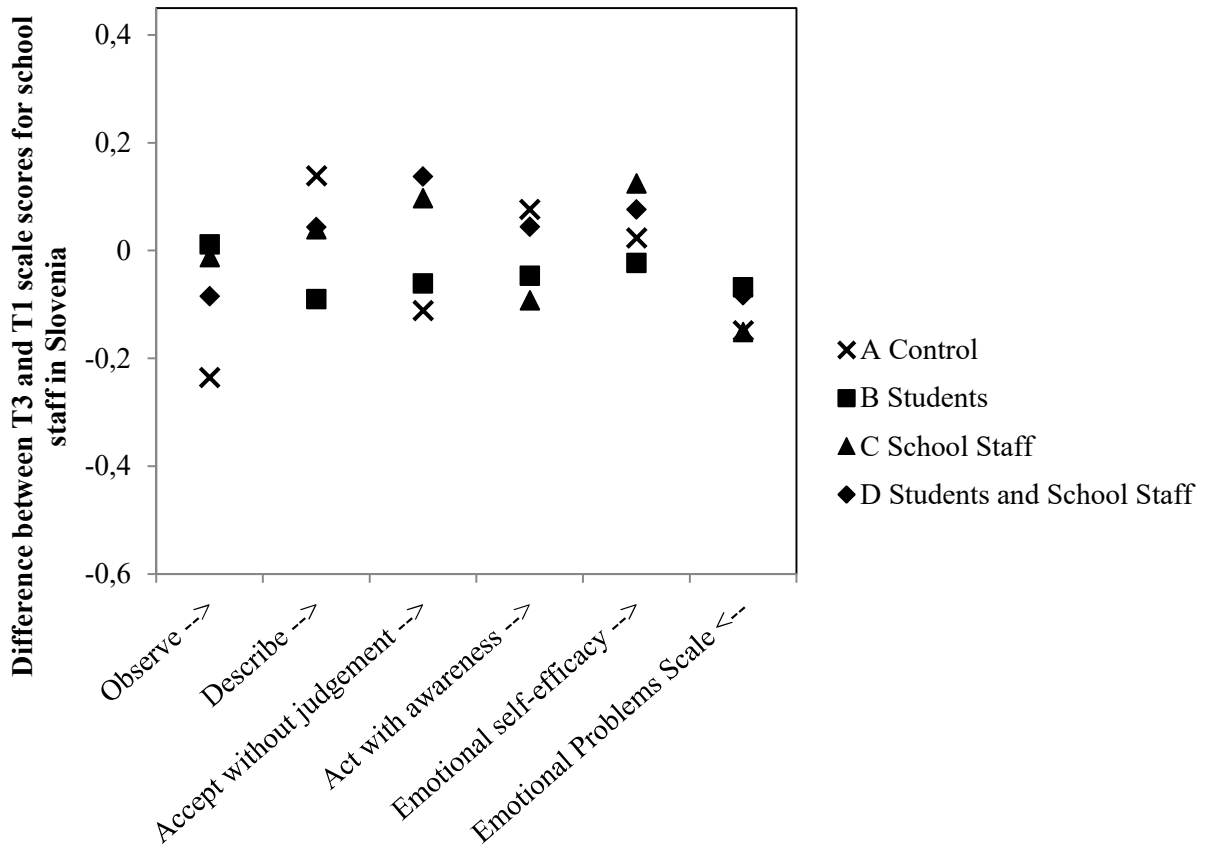


Figure 5.37. Average difference between T3 and T1 per group for the scales assessing the school staff's self-awareness and self-management in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

The average differences between the two points in time for the control group and the three experimental groups and scales that measure school staff's self-awareness and self-management are presented in Figure 5.37. We find a significant effect in the expected direction for the *observe* scale. In conditions B and C, the scale score does not change while in the control group it decreases (B: $t=2.55$, $p=0.012$, $d=0.32$; C: $t=2.43$, $p=0.017$, $d=0.29$). However, the effect sizes are small and no such effect is found for group D that also experienced the student and school staff programmes. Another effect in the expected direction is observed

with the scale *accept without judgement*. School staff from groups C and D show an increase in the scale scores, while in the control group the scale score decreases (C: $t=2.36, p=0.020, d=0.29$; D: $t=3.03, p=0.003, d=0.33$). Effect sizes are small.

5.4.15. Slovenia: School staff's Relationship Skills and Social Awareness

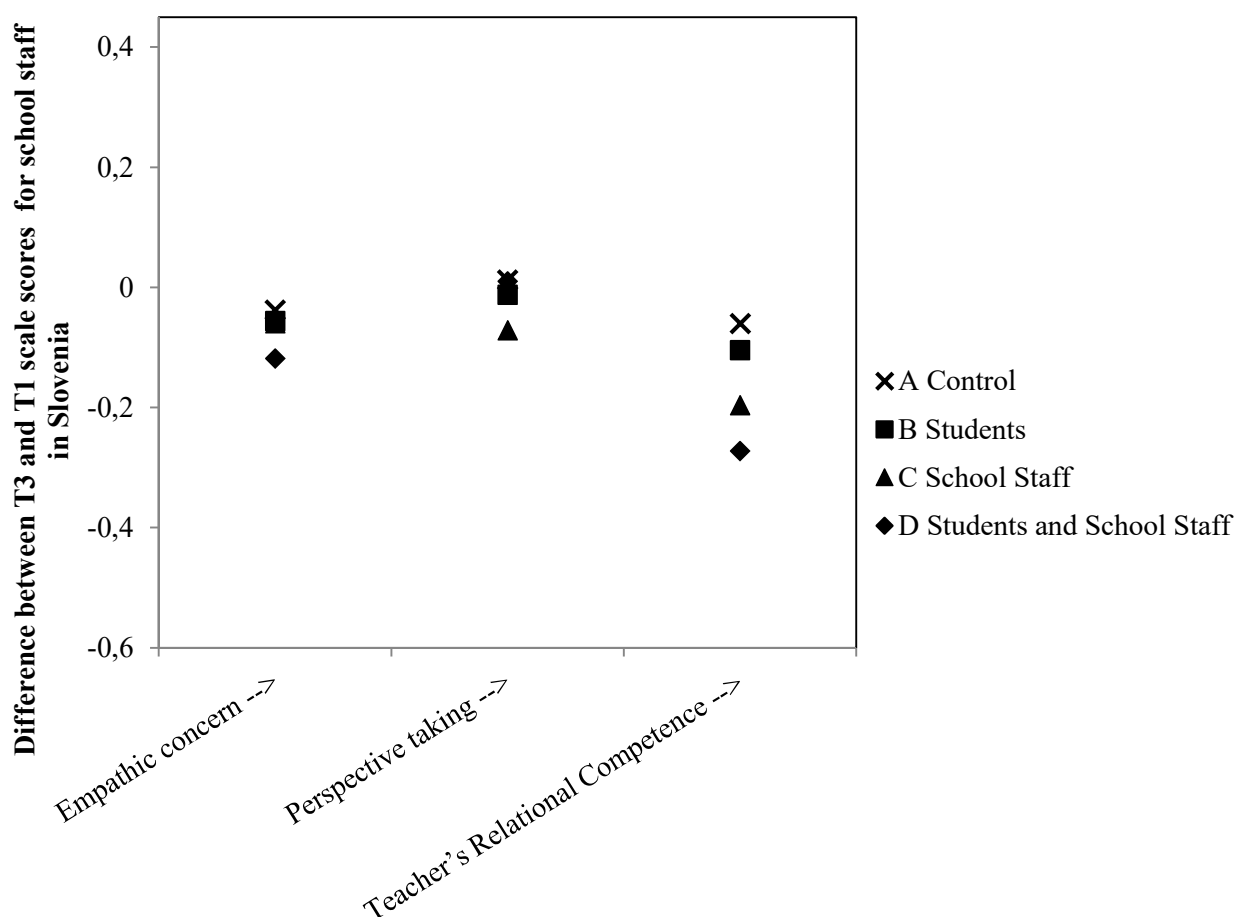


Figure 5.38. Average difference between T3 and T1 per group for the scales assessing the school staff's relationship skills and social awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figure 5.38 presents the average differences between the two points in time for the control group and the three experimental groups and scales that measure relationship skills and social awareness. On the *teacher's relational competence* scale we find an effect in the unexpected direction for group D. While in the control group, the reported relational competence does not change from T1 to T3, it decreases in experimental condition D (trainings for students and school staff).

The difference between group D and the control group is significant, but the effect sizes is small ($t=-3.35$, $p=0.001$, $d=-0.32$). Also, no such difference is found for group C (only staff trainings).

5.4.16. Slovenia: School Staff's Intercultural Competencies/Diversity Awareness

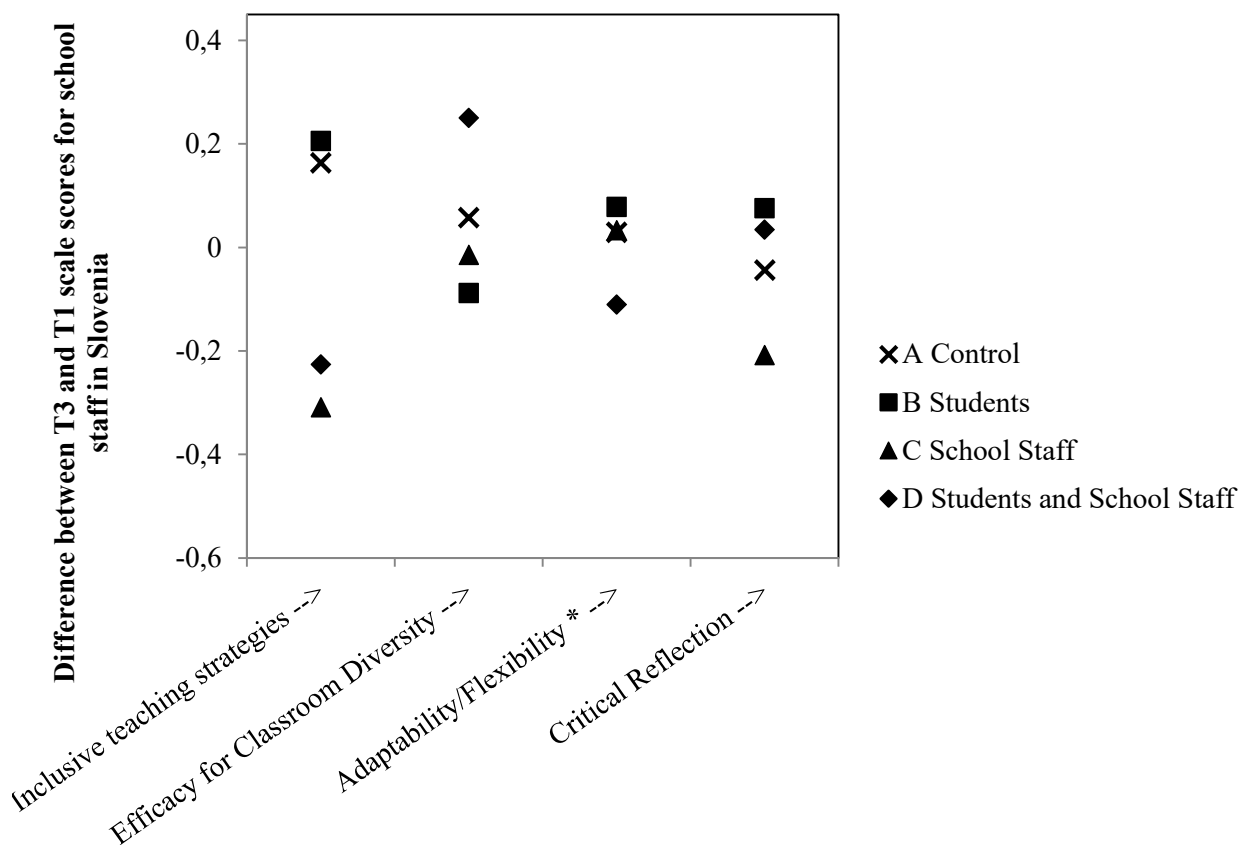


Figure 5.39. Average difference between T3 and T1 per group for the scales assessing the school staff's intercultural competencies/ diversity awareness in Slovenia

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.
 * Reliability of this scale was between .60 and .70 in T1.

Figure 5.39 shows the average differences between the groups for scales that measure school staff's intercultural competencies/ diversity awareness. An effect in the unexpected direction can be observed for *inclusive teaching strategies* for groups C and D. The self-reported implementation of inclusive teaching strategies decreased in experimental groups C and D, while it increased in the control group (C: $t=-3.41$, $p=0.001$, $d=-0.56$; D: $t=-2.37$, $p=0.020$, $d=-0.46$). Effect sizes are medium.

5.4.17. *Summary of Middle-Term Effects*

For analysis of the middle-term effects only results from Croatia and Slovenia were considered, because the response rate in Sweden at T3 was considered too low to draw valid conclusions (see Chapter 4).

Table 5.3. presents an overview over middle-term-effects of the HAND in HAND programmes on students' self-reported social and emotional competencies and intercultural competencies/diversity awareness as well as on their cooperativeness reported by classmates (other report). In Croatia we find only unexpected effects for students, mostly focused in the area of self-management. In Slovenia there are more expected than unexpected effects and these are found for all overarching outcome dimensions. Altogether, there is only one outcome that changed in the expected direction in both experimental groups where the HAND in HAND student programme had been implemented: *observe* in Slovenia. Again, it could be argued that a whole school approach is required to change the school culture and that, therefore, it is possible that a combination of programmes is more effective than only staff programmes for triggering changes in teacher' attitudes, practices and competencies. And, indeed, in Slovenia there are three outcomes where group D developed more positively than the control group, but not group B: *Empathic concern, perspective taking and attitudes towards immigrants*. For all three outcomes a significant effect is also found for group C. This pattern suggests that the school staff programmes might be more effective for changing these student outcomes than the student programme. In Croatia there is no outcome for which positive effects are observed only in experimental group D, but there is indication of a negative effect of a whole-school approach on students' positive identity (there is an additional negative effect of condition D on students' self-control, but because the effect sizes is very small, this is not considered here). However, it is again striking how much the effects vary between school systems and that all effect sizes are small – both for positive expected as well as for negative unexpected effects.

Table 5.3.

Overview over middle-term-effects of the HAND in HAND programmes on students' self-reported social and emotional competencies and intercultural competencies/diversity awareness and as well as their cooperativeness reported by classmates (other report) in different experimental groups.

School System:	Croatia			Slovenia		
Experimental group:	C	B	D	C	B	D
Self awareness						
Positive identity	no	no	-	no	no	no
Observe	no	no	no	+	+	+
Describe	no	no	no	no	no	no
Accept without judgement	no	no	no	no	no	no
Act with awareness	no	no	no	no	-	no
Self management						
Self control	no	no	-	no	-	no
Emotional problems	no	+	no	no	no	no
Aggressiveness	no	+	no	no	no	no
Relationship skills and social awareness						
Empathic concern	no	no	no	+	no	+
Perspective taking	no	no	no	+	no	+
Caring	no	no	no	no	no	no
Cooperation of peers (other report)	no	no	no	no	no	no
Intercultural competence/diversity awareness						
Attitudes towards immigrants	no	no	no	+	no	+
Critical Consciousness	no	no	no	+	no	no

Note: ++ means that the scale score for the respective outcome increased more or decreased less for the respective experimental group as compared to the control group and that the effect size was medium; + means that the scale score for the respective outcome increased more or decreased less for the respective experimental group as compared to the control and that the effect size was small; - means that the scale score for the respective outcome decreased more or increased less for the respective experimental group as compared to the control and that the effect size was small; -- means that the scale score for the respective outcome decreased more or increased less for the respective experimental group as compared to the control and that the effect size was medium; + or – printed in grey means that the effect size was very small; green means that the effect was in the expected direction, light green means that the effect was not necessarily expected but that it is plausible, red means that the effect was in the unexpected direction, light red means that the effect is rather implausible.

Table 5.4. presents an overview over middle-term effects of the HAND in HAND programmes on school staffs' self-reported social and emotional competencies and intercultural competencies/diversity awareness as well as on teachers' relational competence as perceived by students (other report). We find an almost equal number of expected and unexpected effects in Croatia and Slovenia. Again, there is only one outcome that changed in the expected direction in both experimental groups that had participated in the HAND in HAND programmes for school staff: *describe* in Croatia. There is no outcome for which a small or medium or strong positive effect is only observed in group D. There is one outcome for which negative effects are observed across the two experimental groups that had participated in the HAND in HAND programmes for school staff: *inclusive teaching strategies* in Slovenia. There is also one outcome on which only a whole school approach (not the staff trainings alone) appears to have a negative effect: *Teachers' relational competence*. Interestingly, there is an agreement between self- and other-report for this outcome. However, the effect sizes are small for all these effects and, again, there is no scale with effects that are consistent across the two school systems.

Table 5.4.

Overview over middle-term-effects of the HAND in HAND programmes on school staffs' self-reported social and emotional competencies and intercultural competencies/diversity awareness as well as teachers' relational competence as perceived by students (other report).

School System:	Croatia			Slovenia		
Experimental group:	B	C	D	B	C	D
Self awareness						
Observe	no	+	no	+	+	no
Describe	no	+	+	no	no	no
Accept without judgement	++	no	no	no	+	+
Act with awareness	no	no	no	no	no	no
Self management						
Emotional self-efficacy	no	-	no	no	no	no
Emotional problems	no	no	no	no	no	no
Relationship skills and social awareness						
Empathic concern	no	no	-	no	no	no
Perspective taking	no	no	no	no	no	no
Teachers' relational competence (self-report)	no	no	no	no	no	-
Teachers' relational competence (student report)	no	no	+	+	no	-
Intercultural competence/diversity awareness						
Inclusive teaching strategies (self-report)	no	no	no	no	--	--
Self-efficacy for addressing classroom diversity	no	no	no	no	no	no
Adaptability/Flexibility	no	no	no	no	no	no
Critical Reflection	no	-	no	no	no	no

Note: ++ means that the scale score for the respective outcome increased more or decreased less for the respective experimental group as compared to the control group and that the effect size was medium; + means that the scale score for the respective outcome increased more or decreased less for the respective experimental group as compared to the control and that the effect size was small; - means that the scale score for the respective outcome decreased more or increased less for the respective experimental group as compared to the control and that the effect size was small; -- means that the scale score for the respective outcome decreased more or increased less for the respective experimental group as compared to the control and that the effect size was medium; + or – printed in grey means that the effect size was very small; green means that the effect was in the expected direction, light green means that the effect was not necessarily expected but that it is plausible, red

means that the effect was in the unexpected direction, light red means that the effect is rather implausible.

5.5. Participants' view on the HAND in HAND programmes

In addition to the constructs described above, the questionnaires included single items asking participants about their perception of the HAND in HAND programmes. Participants were asked to assess the perceived usefulness of the programmes and also give their opinion about specific aspects of the programmes. These questions were included in the questionnaires only at the second measurement time point, after the programmes had been conducted (with the exception of two questions in the school staff questionnaire that were also administered in T3).

In the following, we start with the presentation of the results for students and continue with the presentation of the results for school staff. In the first part only experimental groups B and D are included in the tables as only these students experienced the student programme, and in the second part only groups C and D are considered as only those participants had experienced the programmes. We conclude with presenting the results of perceived improvements in Social and emotional competencies and intercultural competencies/ diversity awareness by school staff that are presented for all four conditions.

5.5.1. Students

As can be seen from Table 5.5.5. the majority of students in all school systems perceive the HAND in HAND programme as quite or very useful for their everyday life.

Table 5.5.

Students opinion about the usefulness of the HAND in HAND programme for their everyday life by country and condition (%)

	Croatia		Slovenia		Sweden	
	B Students	D Students and School Staff	B Students	D Students and School Staff	B Students	D Students and School Staff
Not useful	10.7	20.7	6.5	2.7	8.9	6.9
Somewhat useful	23.2	29.3	17.7	25.3	11.1	31.0
Quite useful	23.2	25.9	35.5	32	42.2	37.9
Very useful	41.1	24.1	40.3	40	37.8	24.1

The next questions were dealing with different aspects of the HAND in HAND programme and are presented in 5.6. Most of the students in all school systems and conditions rate the theoretical content of the programme as good with the exception of students in condition D in Slovenia where they rate it mostly as fair. The ratings for inner exercises show that most students in Croatia and Sweden, condition B, assess the inner exercises as fair. Most students in Sweden, condition D, rate them as good and most students in Croatia, condition D and Slovenia, both conditions, even rate them as excellent.

The majority of Students in Slovenia and Croatia, both conditions, and Sweden, condition D, rate the physical exercises as good or excellent while almost half of the Swedish students from condition B think they are rather fair. The responses about the training atmosphere are almost equally distributed in Croatia for condition B. In other school systems and conditions students rate the atmosphere more often as good or excellent than poor or fair.

The next question was about trainers' knowledge. With the exception of Swedish students in condition B the majority of students rated the trainers' knowledge as good or excellent. The last aspect students have rated was their trainers' ability to

relate to participants. While the Swedish students in condition B mostly perceived it as fair or good, the majority of students in other groups and school systems rate the trainers' ability to relate to participants as good or excellent.

Table 5.6.

Students' rating of different aspects of the HAND in HAND programme by country and condition (%)

	Croatia		Slovenia		Sweden	
	B Students	D Students and School Staff	B Students	D Students and School Staff	B Students	D Students and School Staff
Theoretical content						
Poor	25.0	8.6	4.8	0.0	21.3	1.7
Fair	23.2	27.6	12.7	16.0	42.6	43.1
Good	30.4	34.5	50.8	45.3	29.8	43.1
Excellent	17.9	27.6	30.2	37.3	6.4	12.1
Inner exercises						
Poor	19.3	8.6	3.2	1.3	21.3	8.8
Fair	29.8	22.4	9.5	10.7	40.4	28.1
Good	28.1	27.6	41.3	40.0	31.9	40.4
Excellent	21.1	41.4	42.9	46.7	6.4	22.8
Physical exercises						
Poor	22.8	10.2	7.9	2.7	21.3	1.7
Fair	19.3	28.8	6.3	6.7	46.8	15.5
Good	28.1	22.0	42.9	42.7	27.7	43.1
Excellent	28.1	39.0	41.3	46.7	4.3	39.7
Atmosphere during the training						
Poor	29.8	11.9	4.8	2.7	19.6	1.7
Fair	19.3	30.5	14.3	16.2	37.0	24.1
Good	24.6	23.7	41.3	36.5	32.6	44.8
Excellent	26.3	32.2	36.5	43.2	10.9	29.3
Trainers' knowledge of the content						
Poor	10.7	8.5	3.2	2.7	15.2	0.0
Fair	17.9	16.9	9.5	8.2	41.3	15.5
Good	25.0	16.9	34.9	42.5	28.3	36.2
Excellent	41.1	55.9	50.8	43.8	13.0	48.3

Trainers' ability to relate to the participants						
Poor	12.3	10.2	3.2	2.7	13.0	3.5
Fair	17.5	20.3	11.1	9.3	43.5	14.0
Good	35.1	20.3	30.2	38.7	34.8	47.4
Excellent	31.6	47.5	52.4	46.7	6.5	35.1

Students were further asked, if they practiced exercises from the HAND in HAND training during a lesson. The results are presented in

Table 5.7. In Croatia and Slovenia, condition B, only 12-15% of students practiced exercises from the HAND in HAND programme during a lesson, in Sweden this percentage is higher (20-24%). Almost half of the students from Slovenia, condition D, reported that they practiced some exercises from the training during a lesson.

Table 5.7.

Students' responses (%) to the question, if any of their teachers ask them to practice exercises from the HAND in HAND training during a lesson

	Croatia		Slovenia		Sweden	
	B Students	D Students and School Staff	B Students	D Students and School Staff	B Students	D Students and School Staff
Yes	12.3	15.3	14.3	43.2	20.8	24.1
No	63.2	52.5	28.6	14.9	50.0	65.5
I am not sure	22.8	32.2	57.1	41.9	27.1	10.3

5.5.2. School staff

The same questions about the HAND in HAND programme were also presented to the school staff. The majority of the school staff in all school systems and conditions expressed that the HAND in HAND programme is not or only somewhat useful for their work. The results are presented in Table 5.8.

Table 5.8.

School staff's rating of the HAND in HAND programme's usefulness for their work by country and condition (%)

	Croatia		Slovenia		Sweden	
	C School Staff	D Students and School Staff	C School Staff	D Students and School Staff	C School Staff	D Students and School Staff
Not useful	52.2	23.8	46.4	31.0	60.0	46.2
Somewhat useful	34.8	42.9	35.7	62.1	40.0	38.5
Quite useful	13.0	28.6	14.3	3.4	0.0	7.7
Very useful	0.0	0.0	0.0	0.0	0.0	7.7

The next set of questions asked about specific aspects of the programme. The results are presented in Table 5.9.: Most of the school staff in all school systems and conditions rate the theoretical content of the programme as good or excellent. The vast majority of the school staff in all school systems and conditions rate the inner exercises of the programme as good or excellent. The same holds true for physical exercises, the atmosphere during the trainings, the trainer's knowledge of the content of programme and the trainer's ability to relate to participants of the content of the programme.

Table 5.9.

School staff's rating of different aspects of the HAND in HAND programme by country and condition (%)

	Croatia		Slovenia		Sweden	
	C School Staff	D Students and School Staff	C School Staff	D Students and School Staff	C School Staff	D Students and School Staff
Theoretical content						
Poor	0.0	0.0	0.0	0.0	0.0	0.0
Fair	0.0	5.0	3.7	0.0	10.0	23.1
Good	47.8	70.0	37.0	46.2	40.0	53.8
Excellent	52.2	25.0	59.3	53.8	50.0	23.1
Inner exercises						
Poor	0.0	5.0	0.0	0.0	0.0	0.0
Fair	4.3	10.0	3.7	0.0	0.0	8.3
Good	21.7	20.0	44.4	29.6	20.0	50.0
Excellent	73.9	65.0	51.9	70.4	80.0	41.7
Physical exercises						
Poor	0.0	5.0	0.0	0.0	0.0	0.0
Fair	4.3	5.0	0.0	0.0	0.0	0.0
Good	13.0	30.0	25.9	30.8	10.0	50.0
Excellent	82.6	60.0	74.1	69.2	90.0	50.0
Atmosphere during the training						
Poor	0.0	0.0	0.0	0.0	0.0	0.0
Fair	0.0	0.0	0.0	0.0	0.0	0.0
Good	0.0	5.3	14.8	15.4	20.0	16.7
Excellent	100.0	94.7	85.2	84.6	80.0	83.3
Trainers' knowledge of the content						
Poor	0.0	0.0	0.0	0.0	0.0	0.0
Fair	0.0	0.0	3.7	0.0	0.0	0.0
Good	8.7	20.0	22.2	11.5	10.0	33.3
Excellent	91.3	80.0	74.1	88.5	90.0	66.7
Trainers' ability to relate to the participants						
Poor	0.0	0.0	0.0	0.0	0.0	0.0
Fair	0.0	0.0	0.0	0.0	0.0	8.3
Good	4.3	20.0	14.8	20.0	10.0	16.7
Excellent	95.7	80.0	85.2	80.0	90.0	75.0

The next three questions were asking school staff, whether they observed any improvements with regard to the social, emotional and/or intercultural competencies of the students in the target class during the past 4 (in T2) or 6 months (in T3), respectively. The results are presented in Table 5.10. for Croatia

and in Table 5.11. for Slovenia separately. Due to the very low response rates in Sweden the table for Sweden is omitted.

In Table 5.10. the results are presented for Croatia. We can see that many members of the school staff did not observe any changes in Social and emotional competencies and intercultural competencies/ diversity awareness in conditions A, B and C between T1 and T2. Most of the observed changes were reported for conditions A and C. If we look at the distribution of responses for changes in T3 we notice that the school staff observed changes in more students than in T2. Surprisingly the school staff from the control condition reports more improvements in students' Social and emotional competencies and intercultural competencies/ diversity awareness compared to the school staff from experimental conditions. The least improvements in students' Social and emotional competencies and intercultural competencies/ diversity awareness were reported by school staff from students only condition (B).

Table 5.10.

Improvement with regard to the social, emotional and/or intercultural competencies of the students in the target class reported in T2 and T3 as perceived by the school staff by condition in Croatia (%)

	Croatia T2				Croatia T3			
	A	B	C	D	A	B	C	D
Social competencies								
No, for none or almost none of the students	37.5	41.2	25.0	11.8	16.7	29.4	12.5	6.7
Yes, for a few students	25.0	41.2	45.0	70.6	22.2	52.9	37.5	53.3
Yes, for about half of the students	18.8	11.8	10.0	17.6	16.7	17.6	31.3	13.3
Yes, for many of the students	18.8	5.9	20.0	0.0	44.4	0.0	18.8	26.7
Yes, for all or almost all of the students	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Emotional competencies								
No, for none or almost none of the students	37.5	35.3	30.0	17.6	16.7	29.4	12.5	13.3
Yes, for a few students	25.0	47.1	35.0	58.8	22.2	58.8	37.5	46.7
Yes, for about half of the students	12.5	11.8	25.0	23.5	11.1	11.8	43.8	20.0
Yes, for many of the students	25.0	5.9	10.0	0.0	50.0	0.0	6.3	20.0
Yes, for all or almost all of the students	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Intercultural competencies								
No, for none or almost none of the students	43.8	52.9	35.0	17.6	22.2	41.2	12.5	26.7
Yes, for a few students	18.8	35.3	40.0	64.7	16.7	47.1	43.8	33.3
Yes, for about half of the students	12.5	5.9	15.0	17.6	22.2	11.8	25.0	13.3
Yes, for many of the students	25.0	5.9	10.0	0.0	38.9	0.0	18.8	26.7
Yes, for all or almost all of the students	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Note: A - control group, B - students, C - school staff, D - students and school staff

The observed improvements with regard to the social, emotional and/or intercultural competencies of the students for Slovenia are presented in Table 5.11. The least improvement from T1 to T2 according to school staff was present for students' intercultural competencies/ diversity awareness in the school staff only (C) condition. Otherwise the reported improvements do not vary considerably between different competencies within conditions. Similar to Croatia, the school staff in Slovenia reported about more improvements half a year after the programme had ended (T3). Many improvements are reported for social and emotional competencies in the control group and for condition D (students and school staff) whereas the intercultural competencies/ diversity awareness changed most in the students only condition (according to the observations of their respective teachers).

Table 5.11.

School staff's observed improvement with regard to the social, emotional and/or intercultural competencies of the students in the target class for T2 and T3 by condition in Slovenia

	Slovenia T2				Slovenia T3			
	A	B	C	D	A	B	C	D
Social competencies								
No, for none or almost none of the students	61.5	53.3	36.4	29.6	33.3	38.1	50.0	11.5
Yes, for a few students	23.1	40.0	45.5	59.3	33.3	33.3	35.0	69.2
Yes, for about half of the students	7.7	3.3	13.6	7.4	16.7	19.0	5.0	11.5
Yes, for many of the students	7.7	3.3	4.5	3.7	13.3	9.5	5.0	7.7
Yes, for all or almost all of the students	0.0	0.0	0.0	0.0	3.3	0.0	5.0	0.0
Emotional competencies								
No, for none or almost none of the students	59.0	43.3	45.5	33.3	33.3	42.9	35.0	11.5
Yes, for a few students	25.6	46.7	36.4	51.9	33.3	28.6	40.0	61.5
Yes, for about half of the students	10.3	6.7	13.6	11.1	16.7	14.3	15.0	15.4
Yes, for many of the students	5.1	3.3	4.5	3.7	13.3	14.3	10.0	11.5
Yes, for all or almost all of the students	0.0	0.0	0.0	0.0	3.3	0.0	0.0	0.0
Intercultural competencies								
No, for none or almost none of the students	69.2	56.7	59.1	40.7	40.0	33.3	50.0	19.2
Yes, for a few students	17.9	33.3	31.8	51.9	40.0	38.1	35.0	61.5
Yes, for about half of the students	7.7	3.3	9.1	0.0	10.0	9.5	10.0	11.5
Yes, for many of the students	5.1	6.7	0.0	7.4	6.7	19.0	5.0	7.7
Yes, for all or almost all of the students	0.0	0.0	0.0	0.0	3.3	0.0	0.0	0.0

Note: A - control group, B - students, C - school staff, D - students and school staff

In Table 5.12. we present results on the school staff's observations, in how far they saw students using elements from the training during school during the past 4 (in T2) or 6 months (in T3). In T2 in Croatia none of the school staff members reported observing students using elements from the training, in Slovenia only few school staff members are certain (most of them in condition C) and in Sweden only school staff from the students only condition reported about their students using elements of the training. For T3 more school staff members in all conditions are certain that they observed students using elements from the training (most of them in Croatia in conditions B and D and Slovenia condition C).

Table 5.12.

School staffs' observation of their students using elements from the training spontaneously during the school day by condition and country in T2 and T3 (%)

	Croatia			Slovenia			Sweden		
	B	C	D	B	C	D	B	C	D
T2									
Yes	0.0	0.0	0.0	7.7	10.5	4.8	20.0	0.0	0.0
No	36.4	0.0	45.5	42.3	31.6	38.1	40.0	25.0	77.8
I am not sure	63.6	0.0	54.5	50.0	52.6	57.1	40.0	25.0	22.2
The students did not participate in a HAND in HAND training	0.0	100.0	0.0	0.0	5.3	0.0	0.0	50.0	0.0
T3									
Yes	30.0	0.0	22.2	5.9	16.7	9.5			
No	10.0	0.0	33.3	29.4	27.8	28.6			
I am not sure	50.0	0.0	44.4	58.8	50.0	61.9			
The students did not participate in a HAND in HAND training	10.0	100.0	0.0	5.9	5.6	0.0			

Note: B - students, C - school staff, D - students and school staff

5.6. Discussion

The HAND in HAND programme was implemented in three different school systems with the goal of building more inclusive classrooms, schools and, ultimately, societies for all by helping students, teachers and other school staff to develop their social and emotional competencies and their intercultural competencies/diversity awareness. Whether this goal was accomplished was evaluated using an experimental design that compares three different experimental groups with one control group. Two of the three experimental groups (groups B and D) had received a student training and two of the three experimental groups (C and D) had received school staff trainings. This design implies that a causal effect of the student programme should show up in groups B and D, a causal effect of the school staff programmes in groups C and D. The programmes might additionally have indirect effects on the respective other stakeholder group (the staff programmes on students and, possibly, also the student programme on school staff). Moreover, there is some indication from previous evaluation research that a “whole-school approach” might be more effective than programmes that address only one group of stakeholders in a

school, because the latter might not be sufficient to change the school culture (see also Chapter 1 in this report). Hence, it was also expected to find stronger effects for group D than for groups B and C respectively. This chapter presents the results of the comparisons of manifest changes in students' and school staff's social and emotional competencies and intercultural competencies/ diversity awareness (for the measurement of change in classroom climate, see Chapter 12) of the three different experimental groups with the control group across three points in time.

After analysing the short- and mid-term programme effects, we find only some of the expected programme effects on students' and school staffs' social and emotional competencies and intercultural competencies/diversity awareness.

For students we observed the following effects:

- In Croatia the student programme alone did not have any consistent¹³ effect on students' social, emotional, and/or intercultural competencies/diversity awareness. In combination with the staff programmes it had a small short-term effect on students' ability to *observe* (which is part of mindfulness). This might suggest that a whole-school approach is more effective for helping students develop this competency. However, the effect is not sustainable over time (until T3) and there are also no other significant middle term effects in Croatia.
- In Slovenia the student programme had a positive short-term effect on students' *attitudes towards immigrants* and a positive middle-term effect on students' ability to *observe* – no matter whether it was used alone or in combination with staff programmes. The whole-school approach had more significant effects: On the short term, schools where student and school

¹³ An effect was considered as consistent when it appeared in all experimental groups where we would predominantly expect effects. For example, if an effect is found in the teacher sample in the group where only teachers had the training but not where teachers and students had the training, this was considered as an inconsistent effect and thus not interpreted. The other way around, if there was an effect in the teacher sample in the group where students and teachers were trained but not in the group where only teachers had the training, this was interpreted as the whole school approach being more effective than trainings for just a single group. In other words, effects were only regarded as consistent if there was an effect in the expected direction for C and D or only D in the teacher sample and for B and D or only D in the student sample. predominantly expect effects. For example, if an effect is found in the teacher sample in the group where only teachers had the training but not where teachers and students had the training, this was considered as an inconsistent effect and thus not interpreted. The other way around, if there was an effect in the teacher sample in the group where students and teachers were trained but not in the group where only teachers had the training, this was interpreted as the whole school approach being more effective than trainings for just a single group. In other words, effects were only regarded as consistent if there was an effect in the expected direction for C and D or only D in the teacher sample and for B and D or only D in the student sample.

staff programmes had all been implemented developed more positively than the control group not only with regard to students' *attitudes towards immigrants*, but also with regard to students' *caring* and their ability to take different perspectives (*perspective taking*). Effects in experimental group D on *attitudes towards immigrants* and *perspective taking* are also found at time-point T3, so they appear to be sustainable over 6 months after the end of the programme. Additionally, middle-term effect in experimental group D were also observed on students' ability to *observe* and students' *empathic concern*. Notably, for *empathic concern*, *perspective taking* and *attitudes towards immigrants* a significant effect is not only found for group D, but also found for group C and not for group B. This pattern suggests that the school staff programmes might have been more effective for changing these student outcomes than the student programme itself. Possibly teacher modelling and/or support for behaviour changes is more important here than the exercises for students themselves.

- In Sweden only short-term effects could be analysed and only one short-term effect was consistent over the two experimental groups where a student programme had been implemented: In these groups students reported a more positive development of *self-control* as compared to the control group. There is further no evidence that a whole-school approach is more effective: All significant effects found in experimental group D are also found in group B.

With regard to effects on school staffs' social and emotional competencies and intercultural competencies/diversity awareness the following effects were observed:

- In Croatia a combination of student and school staff programmes had a short-term effect on school staffs' ability to *observe* (which is one aspect of mindfulness). No such effect is observed for the school staff programmes alone and, also for experimental group D, the effect was not sustained until T3. At T3, however, a difference between experimental group C and the control group was found for the mindfulness dimension *observe* with a more positive development in group C. Moreover, one middle-term effect was found that was significant and consistent across experimental groups C and D for the mindfulness-dimension *describe*.

- In Slovenia a combination of the school staff and student programmes (experimental group D) had an effect on school staff's self-reported *self-efficacy for addressing classroom diversity* and on their *empathic concern*. Yet these effects were not sustained until T3. On the middle-term (6 months after the training) only negative effects were observed in Slovenia: One on *inclusive teaching strategies* in experimental groups C and D and one on *teachers' relational competence* in group D only.
- In Sweden only short-term effects could be analysed and the findings suggest that the HAND in HAND student programme, the school staff programmes and a combination of all three all had an effect on the mindfulness-dimension *observe*.

It should, however, be noted that most of these effects had small effect sizes (Cohen's D between 0.20 und 0.50) and some even very small effect sizes (Cohen's $D < 0.20$). Only very few had medium effect sizes and none of the effects was strong. It is further striking, that the results of this evaluation vary substantially across the three school systems, suggesting effect-heterogeneity at the system level. The heterogeneity may be explained by the fact that different trainers implemented the programme in different school systems. The school systems also applied different sampling strategies. Moreover, characteristics of the three school systems (see e.g. Štremfel, 2020) as well as specific school characteristics (the school samples were small and unrepresentative of the target population for each country) may have played a role. Yet, if the active ingredients of the programmes were really effective as such, then at least some similarity in effect patterns across school systems would have been expected. Instead we found no single effect of the programmes that was consistent across school systems.

Notably, there is one scale for which a number of effects is observed for students as well as for school staff, in different school systems, and at different time points: the scale *observe*. Also for this outcome none of the observed effects are consistent across school systems and not all of them are consistent across experimental conditions within each school system, but, at least, 12 out of 30 possible effects are positive and significant. This scale measures a specific aspect of mindfulness. More specifically, the questions in the questionnaire ask whether participants commonly observe, notice or attend to various stimuli, including internal phenomena (cognitions, bodily sensations) and external phenomena

(sounds, smells). Practising this kind of unjudgemental observing is a central element of the HAND in HAND programmes. It seems that under certain condition the HAND in HAND programmes can support the development of this aspect of mindfulness. Yet, considering the inconsistencies in our results, more research is needed to corroborate this impression.

In addition to the expected effects, described in the previous paragraphs, also several effects in the unexpected direction were observed. Hence, some competencies changed more in a positive direction or less in a negative direction in the control group as compared to the experimental groups. The effect sizes were small for most of these effects and inconsistent across the three school systems. Further, there was only one unexpected effect that was consistent across the two respective experimental groups that had participated in a similar programme within a school system: Teachers reported to use less inclusive teaching strategies at T3 (follow-up) than at T1 in Slovenia. One possible explanation for this effect might be that teachers have started to better understand inclusive teaching through the HAND in HAND programme and have become more self-critical. But it might also be that they really refrained from using such teaching strategies as a consequence of the programme.

5.6.1. Limitations

The data collected for the HAND in HAND programme come with some technical limitations. First, schools were allocated to experimental groups so that all students in the participating class participated together in the same programme and all teachers at the same school and also school leaders of that school together with other school staff. As the experimental manipulation happened at the school-level, it would have been preferable to analyse effects also at this level. However, the sample size at this level was small: altogether only 9 schools/classes per experimental condition and only 3 schools/classes per experimental condition within each school system participated. Therefore, it was not possible to examine school-level effects – at least not separately for each of the three school systems. The small sample size at the system-level also implies that third variables at the school/classroom level could have potentially had a noticeable confounding effect. For example, we know from the interviews about an extremely stressful event that happened to one of the students in one of the schools and that could have potentially influenced the climate in the whole class. Because the sample

size at the school/classroom level was too small, we analysed student-level data. However, for the teachers as well as for the school leader/other school staff the sample sizes were even too small for that. Therefore, we analysed teachers and other school staff together – even though they participated in programmes that differed considerably in length.

A completely different limitation is that in Sweden and Croatia the schools self-selected for the programme (i.e., a convenience sample), holding important implications for the external validity of the results. In Sweden, the drop-out rate during the programme was further relatively high; therefore, special caution should be taken when interpreting the school staff results for T2. Because of further drop out for T3, no analyses could be conducted for school staff. Also for students, a larger decrease in the sample size was present in the control condition where only 10 students responded to the questionnaire, which made statistical comparisons unreliable.

Finally, it is possible that the measures used were not “instruction sensitive” (see Naumann, Hochweber, & Klieme, 2016 for a detailed description of the concept of instructional sensitivity) enough to detect changes. Researchers in the field are only starting to become aware of this issue, therefore there is only few information on the instruction sensitivity of existing scales. We tried to choose instruments that had been used in other experimental studies and for which significant effects had been observed before (which is an indication of instruction sensitivity), but we did not find such instruments for all theoretically relevant dimensions. We applied a mixed-methods approach to avoid this problem. Yet, the evaluation of effects on social, emotional, and intercultural competencies/diversity awareness relied heavily on self-reports.

5.6.2. *Conclusions*

The experimental summative outcome evaluation of the HAND in HAND programme had the aim to investigate whether the programme was effective in terms of triggering changes in social and emotional competencies and/or intercultural competencies/ diversity awareness of students and/or school staff (as well as in classroom climates, but this will be discussed in Chapter 12). The answer appears to be complex: Even though we found several of the expected effects, we also found many unexpected effects. Effect sizes were mostly small or

even very small. Many of the effects were further inconsistent across experimental groups that had participated in a similar programme (groups B and D and groups C and D respectively), but we also did not find a clear evidence for the superiority of the so-called “whole-school condition” (the condition with student and school staff programmes combined). Moreover, there was not a single significant effect that was consistent across school systems. Differences between school systems can have several reasons – among other things that the trainers were different in each system – yet, if the active ingredients were effective as such, then we would have expected at least some overlap in the pattern of effects. Thus, our findings do rather not support a generalizable effectiveness of the programme.

The technical limitations described in the previous section might be one explanation for our mixed findings. In particular, the small sample size at the school level. Schools are complex systems and triggering change in these complex systems might depend on many contextual factors that are impossible to control for in small experimental studies. (It should, however, be noted that the HAND in HAND study was already a large and elaborate study and that it is extremely costly and time-consuming to include a larger sample of schools in an experimental study.) A further possible reason why we found only few consistent effects is that our instruments might not have been suited to detect exactly those changes triggered by the programme – even though we applied a mixed-methods approach. So, it is possible that the programme had effects that we were just not able to identify. It is also possible that the programme did not have consistent positive effects in the expected direction. Maybe it had effects only in some schools and not in others. Maybe a longer programme is needed to bring about more consistent changes in social and emotional competencies and intercultural competencies/diversity awareness. More research is needed to come to a definite conclusion.

5.7. References

IBM Corp. (Released 2013). *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.

Kozina, A., Vidmar, M., & Veldin, M. (2020). Social, Emotional and Intercultural/transcultural Learning in a European Perspective: Core

Concepts of the HAND in HAND Project. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (17-37). Hamburg: Verlag Dr. Kovač.

Maas, C.J.M., & Hox, J.J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology*, 1, 86-92.

Naumann, A., Hochweber, J., & Klieme, E. (2016). A psychometric framework for the evaluation of instructional sensitivity. *Educational Assessment*, 21 (2), 89-101.

Rožman, M., Roczen, N., & Vieluf, S. (2020). Evaluation of the HAND in HAND programme: Results from questionnaire scales. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (157-194). Hamburg: Verlag Dr. Kovač.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>

Štremfel, U. (2020). The embeddedness of social, emotional and intercultural/transcultural learning in European and national educational policies and practices. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (39-58). Hamburg: Verlag Dr. Kovač.

Vieluf, S., Denk, A., Rožman, M., & Roczen, N. (2020). How do the participants evaluate the HAND in HAND programme? Results of semi-structured focus group interviews. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awareness for inclusive school environments across Europe* (195-218). Hamburg: Verlag Dr. Kovač.

Chapter 6:
Changes in intercultural competence/diversity awareness: Results
from a vignette study

Svenja Vieluf, Jan Herlyn, Franziska Zirker, Mojca Rožman,
Nina Roczen

6.1. Introduction

The summative outcome evaluation aimed at assessing whether the Hand in Hand programmes had an effect on the socio-emotional competencies and intercultural competencies/diversity awareness of participating students and school staff (see Chapter 2 of the report). Chapter 5 described the results of analyses of changes in socio-emotional competencies and intercultural competencies/diversity awareness in different experimental groups based on self-report measures and other-report measures. This chapter presents analyses of changes in participants' intercultural competencies/diversity awareness based on their responses to a vignette describing a social situation, which triggers beliefs about the social difference category "migration background".

Vignettes were used to increase the validity of conclusions about programme effects. Self-report measures are subject to several biases and limitations (see also Chapter 3): First, subjects' interpretation of the question can differ from the meaning researchers had intended. Second, the validity of responses will depend on subjects' introspective ability as well as on their correct recall and interpretation of past social events, people's behaviours, thoughts and feelings. This means that subjects with higher socio-emotional competencies and intercultural competencies/diversity awareness will probably be better able to provide accurate descriptions of exactly these competencies. Third, even if the subject is able to assess him*herself accurately, he*she also needs to be willing to answer honestly to the questions in the questionnaire (see e.g., Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Keefer, 2014; Krosnick & Presser, 2010; Paulhus & Vazire, 2007). Moreover, there are individual differences in subject's use of Likert-type response scales: Individual subjects might be more or less prone to response styles such as acquiescence, extreme responding or social desirability (e.g., Jackson & Messick, 1965; Knowles & Condon, 1999; He & van de Vijver, 2013). Other-reports might in some cases be less affected by social desirability-effects. In particular, the perspective of others is important for understanding the subjects' social competencies, as the others can inform better on the impression an individual makes on others and how his*her behaviour is perceived by others. Yet, the accurateness of others in evaluating the behaviour of the subject will also depend on their social competencies. Moreover, questionnaire-based other-reports are just as affected by response style bias as self-reports. Hence, the validity of conclusions on programme effectiveness for changing social and emotional competencies as well as intercultural competencies/diversity awareness based on self-reports or other-reports will be limited.

Vignettes are short stories or scenarios, presented in written form, with pictures, or in a video format (e.g., Hill, 1997; Richman & Mercer, 2002). Participants are typically asked how they interpret the situation described in the scenario and/or how best to respond, how they would personally respond, or how they think a third person would respond (Hughes, 1998). Questions can be closed, requiring subjects to choose among different responses or to indicate their agreement with responses on a Likert-type scale, or they can be open ended (yet, using a Likert-type scale would reduce the advantages over self-report measures). The aim of the vignette-technique is to elicit rich but focused responses from the subjects (e.g., Schoenberg & Ravdal, 2000). It can help elucidate subjects' own interpretations of social situations and meanings as well as their individual belief systems (e.g., Barter & Renold, 2000; Schoenberg & Ravdal, 2000). In contrast to self-reports, introspective abilities and the ability to recall past events have only little influence on the validity of responses. Responses to vignettes can also be affected by social desirability, but this influence is probably smaller than for self-report scales. In particular, including a character in the scenario who breaks social norms can reduce the pressure on subjects to provide socially desirable answers (Barter & Renold, 2000). Also, the relative distance between the vignette and the subject can help addressing sensitive issues and talk about topics that can trigger strong negative emotions, such as mobbing or racism. For example, Barter and Renold (2000) observed commenting on stories about other people's experiences was viewed as less threatening compared to talking about the own experiences and feelings. Vignettes have previously been used successfully to measure intercultural competences/diversity awareness (e.g., Busse & Krause, 2015; Hesse & Göbel, 2007; Rockstuhl, Ang, Ng, Lievens, & Van Dyne, 2015; Schwarzenhal, Juang, Schachner, & van de Vijver, 2019; Whetzel & McDaniel, 2009). Vignettes also have drawbacks. In particular, they only assess interpretations, beliefs and meanings in relation to a very specific situation (Schwarzenhal et al., 2019). But, as the drawbacks are different than those of self- and other-reports, vignettes can be considered an important complement to the former measures and key element in a multi-methods approach.

This chapter starts with a description of the vignettes and the coding system developed. Consequently, the statistical analyses of differences between T1 and T3 in the frequency of response types are briefly described before the results of these analyses are presented and conclusions are drawn.

6.2. The vignette used for assessing diversity orientations at different points in time for the evaluation of the Hand in Hand programmes

The Hand in Hand evaluation used a vignette that had been developed by Schwarzenhal, Juang, Schachner, and van de Vijver (2019) to measure intercultural competence of adolescents. The scenario described in the Student Questionnaire is the following:

A student who has just immigrated to <test country> is new to your class. A month later, you notice that he is still sitting in a corner on his own during school break times. He is sometimes late for lessons and usually has not done his homework. Some classmates give him funny looks but they do not talk to him.

In the School Staff Questionnaire a slightly modified version was included:

A student who has just immigrated to <test country> is new to a class you are teaching. You notice that he is always sitting in a corner on his own during school break times. He is sometimes late for lessons and usually has not done his homework. Some classmates sometimes give him funny looks but they do not talk to him.

The description of the scenario was followed by open-ended questions assessing the subjects' interpretation of the incident and suggestions how best to respond. Students were asked:

1. Why is the student behaving in that way? (Please write at least two sentences!)
2. Why are your classmates behaving in this way? (Please write at least two sentences!)
3. What would you do if you were observing this situation? (Please write at least two sentences!)

School staff was asked:

4. What do you think, why is the student behaving in that way? (Please write at least two sentences!)
5. What do you think, why are his classmates behaving in this way? (Please write at least two sentences!)
6. What would you do if you were observing this situation in your school? (Please write at least two sentences!)

6.2. Development of a coding scheme

Schwarzenthal and others (2019) used three coding categories for the first two questions (which asked about reasons for the behaviour of the new student and of his classmates):

1. the degree to which subjects considered cultural influences in their responses (3-point Likert scale)
2. the degree to which subjects suspended judgment of the people involved when interpreting the scenario (3-point Likert scale)
3. the degree to which subjects considered alternative explanations when interpreting the behaviour in the incidents (3-point Likert scale)

They further used one coding category for the third question (which asked how the students would behave if they were in a similar situation):

4. To what extent does the subject's intended behaviour solve the situation effectively, i.e., does he/she find a solution that considers both parties' interests?

Because we used the vignette not only with students, as Schwarzenthal and others (2019) had, but also with school staff members, because we found no variance for some of the coding categories Schwarzenthal and others (2019) had developed in our data, and because the Hand in Hand team had decided to shift the focus from assessing intercultural competence to assessing diversity awareness (see Chapter 3), we developed a new coding scheme. This was done by the first author (master coder) in cooperation with Franziska Zirker and Jan Herlyn (reliability coders) and based on recommendations given by Syed and Nelson (2015). More specifically, we used a theoretically driven inductive approach for the development of coding categories: The coding system was generated inductively, but we drew from theoretical perspectives when identifying and naming themes.

First, all three coders reviewed the material and together we developed lists of initial codes. Based on theoretical perspectives on diversity and inclusion, we then searched for themes in the codes and, cooperatively, developed a dynamic working coding manual. We used this coding manual for coding the whole material. The unit of analysis was each half sentence. We decided for a rather large number of codes, because we wanted to capture even nuanced changes in the perspectives of students and school staff who had participated in a Hand in Hand programme. Responses of students and school staff were coded separately. The material was split in a way that every answer was coded by the first author

plus at least one of the two other coders. We did continuous coding checks throughout the coding process to prevent coder drift. We also discussed open questions and disagreements and, based on these discussions, continuously refined the manual. Related to the dynamic code-development procedure, four basic codes/ sub-categories for coding responses of students to questions 1 and 2 were identified at a later stage of the coding process. These were then applied to the whole material in a second round of coding only by one of the coders, because the other two were not available anymore. To assure quality of these codes, the master coder coded the material twice with 2 months in between and carefully reviewed all disagreements between the first and the second coding.

6.2.1 Coding Responses to Questions 1 and 2

Questions 1 and 2 asked the subjects about reasons for the behaviour of the new student and his classmates described in the scenario, i.e. reasons why the new student does not comply with the school rules, why he is alone, and why the classmates give him funny looks, but do not talk to him. Coders classified the causes mentioned by subjects. These basic codes/ sub-categories were then summarized to main codes which reflect theoretical discourses about diversity reflected in the interpretations and meanings of subjects. The coding scheme for students is shown in Table 6.1. That for school staff is shown in Table 6.2.

The content analysis of student's responses to questions 1 and 2 resulted in 12 basic codes or subcategories (see Table 6.1.). These were summarized to four main categories: (1) Sub-categories in white were summarized to the main category "situational and individual explanations". These are units of meaning referring to situational and individual factors, including emotions, to explain the behaviour of the new student and/or his classmates. (2) The sub-category in grey represents the main category "devaluated otherness". These are units of meaning referring to an otherness of the new student which is devaluated through using the term 'strange' to explain the behaviour of the new student and his classmates. (3) The sub-category in light blue represents the main category "group-based otherness". These are units of meaning referring to an otherness of the new student which is assumed to be group related (often a national heritage, race or religion different from that of a majority is mentioned in these responses) and described in neutral terms. (4) The sub-category in medium blue represents the main category "cultural influences". These are units of meaning referring to a language barrier or differences in norms, values and behavioural expectations between the environment the new student has left behind and the new environment. (5) The sub-categories in darker blue represent the category "social exclusion". These are

units of meaning referring not to an otherness of the new student, but to an ascription of otherness to the new student by the classmates as roots of their exclusive behaviour or referring to exclusive institutional settings and practices.

Table 6.1

Coding Scheme for Responses of Students to Questions 1 and 2

Basic code/ sub-category	Rules	Examples
Being new	The move and change of school of the new student as well as group dynamics in the new school – which are explicitly described in the scenario itself – are mentioned as an explanation for the behaviour of the new student and/or his classmates.	<p><i>“He is new at the school.”</i></p> <p><i>“He has no friends.”</i></p> <p><i>“He feels excluded.”</i></p> <p><i>“Because they don’t know him.”</i></p> <p><i>“Because people usually react this way to new people.”</i></p> <p><i>“Because he withdraws and does not speak to anybody.”</i></p>
Violation of school norms	A violation of school norms by the new student, i.e. coming too late and not doing his homework – which was explicitly described in the scenario – are mentioned as an explanation for the behaviour of the classmates.	<p><i>“He forgets his exercise books at home and doesn’t do schoolwork.”</i></p> <p><i>“Because he is comes late to school and doesn’t do his homework.”</i></p> <p><i>“They don’t know how to react, because it seems like he does not care about school.”</i></p> <p><i>“Because he does not do, what he is supposed to do.”</i></p>
Emotions	Emotions of the subjects – in particular anxiety, shame, and sadness – are mentioned as an explanation for the behaviour of the new student and/or his classmates.	<p><i>“He is afraid.”</i></p> <p><i>“He is ashamed.”</i></p> <p><i>“He is unhappy.”</i></p> <p><i>“Maybe they are afraid of him.”</i></p> <p><i>“They are afraid to approach him.”</i></p>
Individual characteristics	Individual characteristics of the new student – in particular lack of social competencies, lack of cognitive competencies, lack of learning	<p><i>“He is shy.”</i></p> <p><i>“He is insecure.”</i></p>

of the new student	motivation – are mentioned as an explanation for the behaviour of the new student and/or his classmates.	<p><i>„Maybe he wants to become friends with the others but lacks the courage.”</i></p> <p><i>“It is difficult for him to adjust to the new environment.”</i></p> <p><i>„Or maybe he is a bad student.”</i></p>
Characteristics of the new students’ family	Individual characteristics of the new student – in particular a poor quality of child-parent relations or a lack of support with schoolwork at home – are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.	<p><i>“He has problems at home.”</i></p> <p><i>“He has a poor relation with his parents.”</i></p> <p><i>„Probably he doesn’t do his homework, because he has nobody at home who can help him.”</i></p>
Individual characteristics of the classmates	Individual characteristics of the classmates – in particular lack of social competencies – are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.	<p><i>„I think they behave this way, because they don’t know how else to react.”</i></p> <p><i>„Maybe they don’t know how to approach him and how to become friends with him.”</i></p> <p><i>„They are insecure“</i></p> <p><i>„They don’t give a shit about others. They don’t care.”</i></p>
Strangeness of the new student	An argued strangeness of the new student is mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates. It remains open, whether he as an individual is considered to be “strange” or whether his status as an immigrant makes him strange in their eyes.	<p><i>“Maybe they think, he is strange.”</i></p> <p><i>“He is strange.”</i></p> <p><i>“They think he behaves in a strange way.”</i></p>
Otherness	An otherness of the student is mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates. Either students argued that he is	<p><i>“Because he is black.”</i></p> <p><i>“Because he belongs to a different race and has no friends, or because he is in a difficult economic situation.”</i></p>

	<p>“different” from the others or they explicitly mentioned a social group membership of the new student as an explanation for the behaviour of his classmates.</p>	<p><i>“Because he is maybe different.”</i></p> <p><i>“They don’t want to talk to him, because he is an immigrant.”</i></p> <p><i>“He is different from how people usually are.”</i></p>
Language barrier	<p>An assumed language barrier between the new student and all others is mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.</p>	<p><i>“Assuming that the student does not know <test language>, it is probably difficult for him to communicate with the other students and to understand the homework instructions.”</i></p> <p><i>“They cannot find a common language with him.”</i></p>
Cultural differences	<p>Cultural differences between the new student and his classmates are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.</p>	<p><i>“Yet, he has not gotten used to the way we live here, to our culture.”</i></p> <p><i>“He behaves this way, because he is not yet used to the new environment. The teaching, the behaviour of people and the rules are different.”</i></p> <p><i>“Maybe in his country it was less strict with the homework.”</i></p>
Ascription of otherness by the classmates	<p>An ascription of otherness by the classmates, prejudices, xenophobia or racism of the classmates are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.</p>	<p><i>„Maybe they think he is different.”</i></p> <p><i>“They have prejudices.”</i></p> <p><i>“In my class some students do not respect asylum seekers.”</i></p> <p><i>“Because the person is from a different country, it is also possible that people watch her in a different way for racist reasons (which is very wrong!).”</i></p> <p><i>“Probably because they heard only bad things about immigrants on TV and this shapes their perception of the boy.”</i></p>
Institutional settings and practices	<p>Institutional settings and practices are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.</p>	<p><i>„Maybe the school in <test country> does not provide him adequate access.”</i></p> <p><i>“If someone has newly arrived in a new country from another country, then the teachers should adapt the homework to that person.”</i></p>

The content analysis of school staff's responses to questions 1 and 2 resulted in nine basic codes or sub-categories, which were partly similar to those developed for coding students' responses. The sub-categories were summarized to four main categories: (1) Sub-categories in white were summarized to the main category "situational and individual explanations". These are units of meaning referring to situational and individual factors, including emotions, to explain the behaviour of the new student and/or his classmates. This category also includes referrals to specific stressors that can be caused by migration per se. (2) The sub-category in **medium blue** represents the main category "cultural influences". These are units of meaning referring to a language barrier or differences in norms, values and behavioural expectations between the environment the new student has left behind and the new environment. (4) The sub-categories in **darker blue** represent the category "social exclusion". These are units of meaning referring to an ascription of otherness to the new student by the classmates as roots of exclusive behaviour or exclusive institutional settings and practices.

Table 6.2

Coding scheme for coding responses of school staff to questions 1 and 2

Categories	Rules	Examples
Individual characteristics of the new student	Individual characteristics of the new student – in particular lack of social competencies, lack of cognitive competencies, lack of learning motivation – are mentioned as an explanation for the behaviour of the new student and/or his classmates.	<p><i>"He may be shy."</i></p> <p><i>"It is difficult for him to make contact with new people."</i></p> <p><i>"I think it is difficult for him to understand new routines."</i></p> <p><i>"Unwillingness and inability to accept challenges to realize something."</i></p> <p><i>"The student does not know the rules."</i></p>
Characteristics of the new students' family	Individual characteristics of the new student – in particular a poor quality of child-parent relations or a lack of support with schoolwork at home – are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.	<p><i>"Maybe the student has problems at home."</i></p> <p><i>"The student might not get support at home."</i></p> <p><i>"Probably he doesn't do his homework because nobody controls it. And he is probably late at school, because he needs</i></p>

		<i>to get up by himself and therefore often oversleeps.</i>
Characteristics of the new students' classmates	Individual characteristics of the classmates – in particular lack of social competencies – are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.	<p><i>"It is also difficult for them to connect with new people."</i></p> <p><i>"Shyness and insecurity on both sides."</i></p> <p><i>"They don't know what to talk about and how to communicate with him."</i></p> <p><i>"Some don't care. They lack empathy."</i></p>
Group dynamics	Group dynamics, i.e. social-psychological phenomena and principles that occur in groups, are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates. In particular, social exclusion and fear of exclusion as forces affecting the new students' and/or the classmates' behaviour.	<p><i>"The student experiences exclusion."</i></p> <p><i>„Because he does not have a relation with his classmates, he might come too late to school on purpose, because school is something stressful and negative for him."</i></p> <p><i>"The class may be close-knit, homogeneous and together and they experience that a new student comes and is disturbing this."</i></p> <p><i>"Because he does not feel accepted and thinks that he would be rejected if he tried to communicate. Similar, however, the classmates themselves think that the group would exclude them."</i></p>
(Post-)migration stress	School staff member assumes that the situation described in the scenario was, i.a., due to negative effects of migration or post-migration stress experienced by the new student.	<p><i>"He is homesick."</i></p> <p><i>"He might have experienced trauma while he was fleeing or in his home country, he might have lost family members, which caused sleeping problems and poor health."</i></p> <p><i>„Worries whether he will be allowed to stay in <test country>."</i></p>
Otherness	An otherness of the student is mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.	<p><i>"The student is different from them."</i></p> <p><i>"He came from somewhere else."</i></p> <p><i>"Maybe his appearance plays a role (clothing, hair)."</i></p>
Language barrier	An assumed language barrier between the new student and all others is mentioned as an	<i>"The student does not know <test language> and that makes it difficult to communicate."</i>

	<p>explanation for the behaviour of the new student and/ or for the behaviour of his classmates. (Please note: When the language barrier was addressed as an institutional barrier and the monolingual practice of the school was questioned, “a language barrier” was coded “0” and “institutional routines and practices” was coded “1”.</p>	<p><i>“Probably the student has language difficulties and does not understand much.”</i></p> <p><i>“They don’t understand him.”</i></p> <p><i>“Some don’t feel comfortable because they don’t know the language of the new classmate and it is difficult for them to talk with him.”</i></p>
<p>Cultural differences between the new student and his classmates</p>	<p>Cultural differences between the new student and his classmates are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.</p>	<p><i>“His environment has a culture that is different from that in <test country>.”</i></p> <p><i>“Everything is new and there are many new rules that are different from the rules in the environment where he lived before and this makes him feel uncomfortable.”</i></p> <p><i>“Maybe he does not understand the social codes.”</i></p> <p><i>“The student does not adhere to the same social norms that the others are used to.”</i></p>
<p>Difficulties of the classmates to accept an assumed otherness of the new student</p>	<p>School staff member assumes that the situation described in the scenario was, i.a., due to difficulties of the classmates to accept an assumed otherness of the new student. Often school staff members suggested that the classmates might have prejudices or experience xenophobia.</p>	<p><i>“Because they think he is different.”</i></p> <p><i>“Many have not learned to deal with differences.”</i></p> <p><i>“The students have fear of the unknown and of difference.”</i></p> <p><i>“Because he comes from a different country it is difficult for them to accept him and approach him.”</i></p> <p><i>“They cannot accept diversity.”</i></p> <p><i>“They have prejudices against immigrants.”</i></p> <p><i>“This could also be due to racist tendencies or pure racism in the class.”</i></p>
<p>Institutional settings and practices</p>	<p>Institutional settings and practices are mentioned as an explanation for the behaviour of the new student and/ or for the behaviour of his classmates.</p>	<p><i>“The adults in the school have not taken responsibility.”</i></p> <p><i>“He has not received the guidance and support in school that he would have needed.”</i></p> <p><i>“The teachers would need to inform the students about the arrival of a new classmate and guide them to cooperate.”</i></p> <p><i>“He experiences linguistic exclusion.”</i></p>

6.2.2. Coding Responses to Question 3

Question 3 asked the subjects how they would react if they were observing the situation described in the scenario. Coders classified the type of intervention suggested by the subjects. These basic codes/ sub-categories were then summarized to main codes. These were different for students and teachers. Similar to the coding scheme developed by Schwarzenhal et al. (2019) we summarized the basic codes/ subcategories found in students' responses to main categories reflecting the quality of proposed interventions. For teacher, however, we summarized the basic codes/ subcategories depending on their inclusiveness. The coding scheme for students is shown in Table 6.3. That for school staff is shown in Table 6.4.

The content analysis of students' responses to questions 3 resulted in eight basic codes or subcategories shown in Table 6.3. These were summarized to three main categories, similar to the coding scheme developed by Schwarzenhal et al. (2019): Subcategories in white are suggestions that do not solve the situation effectively. Subcategories in light blue are suggestions for interventions that could solve the situation effectively, but do not explicitly consider the perspective of the subjects involved in the situation. The subcategory in medium blue summarizes suggestions for interventions that explicitly consider the perspective of the new student.

Table 6.3

Coding scheme for coding responses of students to question 3

Basic code/ sub-category	Rules	Examples
No intervention	Student says that he*she would not intervene in the situation described in the scenario.	<p><i>“Nothing, because I don’t care.”</i></p> <p><i>“I would probably not talk to him, because I am shy – even though I know how he feels.”</i></p> <p><i>“I would probably feel uncomfortable too, and go away.”</i></p> <p><i>“I would look at him funny, just like everyone else.”</i></p> <p><i>“I would just watch.”</i></p> <p><i>“I would not stick my nose into things that are not my business.”</i></p>
Unsure	Student says that he*she does not know, what he*she would do.	<p><i>“I don’t know.”</i></p> <p><i>“No idea”.</i></p> <p><i>“I don’t know, because I have never been in this situation and I cannot imagine what I would do if I was.”</i></p>
Being friendly	Student says that he*she would be friendly (e.g. smile or greet), but would not actively approach the student.	<p><i>“If the student would try to join us, I would not drive him out.”</i></p> <p><i>“I would say “Hello” to him in the hallway.”</i></p> <p><i>“I would not look at him funny, and I would not make fun of him. I would just greet him.”</i></p> <p><i>“I would try to be kind to him (if he was kind to me).”</i></p>
Ask adult for help	Student says that he*she would ask an adult (mostly a school staff member) for help.	<p><i>„I would talk to a teacher.”</i></p> <p><i>“I would call a teacher. I would call the psychologist.”</i></p> <p><i>“Talk to my parents.”</i></p> <p><i>“Get an adult.”</i></p>

<p>Help</p>	<p>Student says that he*she would try to help the student.</p>	<p><i>“I would help him.”</i></p> <p><i>“I would help him to be on time for the lessons.”</i></p> <p><i>“I would help him with his homework and with the language, etc.”</i></p> <p><i>“I would talk with him about what needs to be done for school, and ask him whether he needs help with it.”</i></p>
<p>Connect</p>	<p>Student says that he*she would try to build a connection with the student.</p>	<p><i>“I would approach him.”</i></p> <p><i>“I would talk with him.”</i></p> <p><i>“I would be with him.”</i></p> <p><i>“I would offer him to become friends.”</i></p> <p><i>“I would try to become friends with him.”</i></p>
<p>Address group dynamics in the class</p>	<p>Student says that he*she would try to convince classmates to better integrate the new student.</p>	<p><i>“I would invite him to play with us.”</i></p> <p><i>“I would acquaint him with the others.”</i></p> <p><i>“I would convince the others to approach him.”</i></p> <p><i>“I would tell the others to be nicer with him.”</i></p> <p><i>“I would come to his defence.”</i></p>
<p>Ask the new student</p>	<p>Student says that he*she would ask the student how he experiences the situation and what he needs</p>	<p><i>„I would ask him, how he was doing and whether he needed something.”</i></p> <p><i>“I would go up to him and ask him what the matter was and listen to him.”</i></p> <p><i>“I would ask him why he was withdrawing.”</i></p> <p><i>“I would talk to the student and try to understand why he was behaving like this.”</i></p> <p><i>„I would ask him, how everything came about (his immigration, etc.) and how he feels now, because I am a very curious person.”</i></p>

The content analysis of school staff's responses to question 3 resulted in eight basic codes or sub-categories, shown in Table 6.4. These were summarized to four main categories: Subcategories in white are subcategories whose inclusiveness is not assessable. Subcategories in light blue reflect an argumentation pattern that potentially excludes the new student from the class community by marking him as different. Subcategories in medium blue include suggested reactions that can be considered integrative. Finally, the subcategories marked in darker blue encompass inclusive and participative suggestions to intervene as a teacher in the situation described in the scenario.

Table 6.4

Coding scheme for coding responses of school staff members to question 3

Basic code/ sub-category	Rules	Examples
Unspecific intervention	School staff member says he*she would do something to improve the situation, but does not specify what.	<i>„Act and try to integrate the student in the classroom.“</i>
Collect more information first	School staff members suggest to first talk to different stakeholder to achieve a better understanding of the situation before acting.	<i>„Talk to the student and try to identify causes for the situation. I would also talk to single other students in the class to see how they perceive the situation.“</i>
Seek advice from experts	Members of the school staff suggest to consult experts for social and emotional learning and/ or diversity, e.g. student health team, pedagogues, social workers, school psychologist, external advice centres.	<i>“I would consult pedagogues.”</i> <i>“Cooperation with the advice service.”</i> <i>“Talk to the student health team, how we should proceed.”</i>
Intervention that exposes the new	School staff members suggest an intervention that would imply classifying the new student as foreign	<i>„I would introduce the student, his home country, culture and situation and that he needs support.“</i>

student and addresses him as the (cultural) “other”	and different (“othering”). Typical for this category are responses suggesting that the student should introduce himself and talk about his heritage culture and/or migration experiences in front of the class. The category also includes suggestions to exclude the student from the class and send him to a specific preparation class for immigrants instead.	<p>„I would also talk to the new student and incite him to give us an understanding of his customs etc.”</p> <p>„It is common practice that students who lack sufficient knowledge of the <language of instruction> visit a specific class for immigrants. It is very likely that a newly immigrated student belongs in this group.”</p>
Compensatory intervention	School staff members suggest an intervention that solely aims at compensating ascribed deficits of the new student, such as a lack of competencies in the language of instruction, a lack of subject competencies, or a lack of knowledge about school rules etc.	<p>„Learning of the language of instruction should be encouraged (in addition to an official preparatory class for immigrants), by addressing the student with simple sentences accompanied by descriptive gestures (e.g.: "Have you eaten yet?" accompanied by a gesture illustrating eating).”</p>
Intervention that supports integration of the new student	School staff members suggest an intervention that either addresses both sides (the new student and his classmates), or only the classmates of the new student, and that has the aim to support the new students’ acceptance by the group and integration into the group. Yet, the intervention potentially exposes the new student as being different, e.g. particularly needy or a person at risk and, thus, reflects a two-group theory that is characteristic of integrative approaches to address diversity. Typical for this category is choosing a mentor amongst the majority students who helps the new student, or involving majority students in a conversation that has the aim to promote understanding for the specific situation of the new student.	<p>„Maybe someone can take him under his wing.”</p> <p>„I would talk with the class how they would feel if they would move to a new environment, to a new school.”</p> <p>„I would talk with the student who has immigrated to <test country> and I would support him. I would advise the other students to accept the new arrival and to help him, the way they might also one day need help.”</p> <p>„I would talk with the students and explain them his situation. I would try to awaken their interest and to help them empathize and understand the situation of the student”.</p>
Inclusive intervention	School staff members suggest an intervention that addresses the whole class without exposing the new student as being different or particularly needy. It treats the situation as a problem of the whole group, not as a problem of	<p>„I would pay more attention to social games in the classroom, students connect through these games and realize that they are all in a similar situation.”</p>

	<p>the new student. Typical for this category are suggestions to use cooperative learning in the classroom or to implement social learning exercises and games with all students. Additionally, interventions addressing institutional barriers for the immigrant student, such as the monolingual habitus of schools, also fall into this category.</p>	<p><i>„I would use peer learning so that they learn to work with everybody and so that nobody stays left out.“</i></p> <p><i>„If the language is a barrier, then I would try to find other ways and possibilities for him so that he can follow the class and be together with his classmates.“</i></p>
<p>Participatory intervention</p>	<p>School staff members suggests to involve students in the process of finding a strategy for improving the situation in the classroom</p>	<p><i>„I would function as a bridge of communication between the group and the individual, but in a way that they find a solution by themselves.“</i></p> <p><i>„I would sit down with the student and try to find out what the reasons are, and how we can try together to find a solution for this.“</i></p> <p><i>„I would create a ,we-tank‘/a cooperative working group in which all are involved.“</i></p>

6.2.3. Inter-rater reliability

To index inter-rater reliability we used percentage agreement and Kappa. Results for each of the categories are shown in Tables 6.5., 6.6., 6.7. and 6.8. Table 6.5 shows that, for all categories of responses of students to questions 1 and 2, the two reliability coders (Coder 2 and Coder 3) agreed in more than 90% of the cases with the master coder (Coder 1). Cohen’s Kappa is also above 0.60 for all categories, indicating substantial strength of agreement, and in about half of the cases even almost perfect strength of agreement is observed (Cohen’s Kappa > .81). These results confirm high reliability of the ratings.¹⁴

Table 6.5

Inter-rater reliability for coding students’ responses to questions 1 and 2

¹⁴ Please note that four categories were identified in the material at a later stage of the coding process. These were applied to the material in a second round of coding. For this round the reliability coders were not available anymore, so that no information on inter-rater reliability is available. To assure quality of these codes, the master coder coded the material twice with 2 months in between and carefully reviewed all disagreements between the first and the second coding.

Coder number:	1 & 2	1 & 3	1 & 2	1 & 3	1 & 2	1 & 3
Basic code/ sub-category	n		Percentage agreement		Cohen's Kappa	
Being new	875	937	92%	95%	.80	.86
Violation of school norms	-	-	-	-	-	-
Individual characteristics of the new student	876	939	88%	91%	.65	.72
Characteristics of the new students' family	-	-	-	-	-	-
Individual characteristics of the classmates	248	248	93%	94%	.81	.83
Emotions	-	-	-	-	-	-
Strangeness of the new student	-	-	-	-	-	-
Otherness of the new student	628	691	89%	90%	.71	.71
Language barrier	-	-	-	-	-	-
Cultural differences	878	942	99%	99%	.74	.74
Ascription of otherness by the classmates	628	691	95%	94%	.70	.71
Lack of institutional support	875	937	100%	99%	1.00	.86

Table 6.6 shows that, for all categories of responses of school staff to questions 1 and 2, the two coders (Coder 1, the master coder, and Coder 2, one of the reliability coders) agreed in more than 90% of the cases. Cohen's Kappa is also above 0.60 for all categories, indicating substantial, and in many cases even

almost perfect strength of agreement (Cohen's Kappa > .81). These results confirm high reliability of the ratings.¹⁵

Table 6.6

Inter-rater reliability for coding school staff members' responses to questions 1 and 2

Coder number:	1 & 2	1 & 2	1 & 2
Basic code/ sub-category	n	Percentage agreement	Cohen's Kappa
Individual characteristics of the new student	444	91%	.78
Characteristics of the new students' family	442	99%	.92
Characteristics of the new students' classmates	444	91%	.81
Group dynamics	443	91%	.82
(Post-)migration stress	441	95%	.63
Otherness	442	-	-
Language barrier	446	96%	.91
Cultural differences between the new student and his classmates	445	96%	.79
Difficulties of the classmates to accept a presumed otherness of the new student	442	92%	.80
Institutional routines and practices	442	98%	.78

¹⁵ Please note that one category ("otherness") were identified in the material at a later stage of the coding process. This was applied to the material in a second round of coding. For this round the reliability coders were not available anymore, so that no information on inter-rater reliability is available. To assure quality of these codes, the master coder coded the material twice with 2 months in between and carefully reviewed all disagreements between the first and the second coding.

Table 6.7 shows that for all categories of responses of students to question 3 the agreement between the master coder (Coder 1) and each of the reliability coders (Coder 2 and Coder 3, who both coded only part of the material) is 90% or above. Cohen's Kappa is also above 0.60 for all categories, indicating substantial, and in many cases even almost perfect strength of agreement (Cohen's Kappa > .81). These results confirm high reliability of the ratings.

Table 6.7

Inter-rater reliability for coding school students' responses to question 3

Coder number:	1 & 2	1 & 3	1 & 2	1 & 3	1 & 2	1 & 3
Basic code/ sub-category	n		Percentage agreement		Cohen's Kappa	
No intervention	864	935	95%	100%	.91	1.00
Unsure	865	936	99%	99%	.95	.95
Friendly, but no active intervention	865	935	99%	99%	.63	.87
Ask adult for help	865	936	99%	99%	.94	.95
Offer help	865	936	97%	97%	.93	.93
Connect	865	936	92%	93%	.84	.85
Address group dynamics in the class	865	936	97%	97%	.89	.90
Ask the new student what he would like	865	936	92%	91%	.71	.68

Table 6.8 shows that, for all categories of responses of school staff to question 3, the two coders (Coder 1, the master coder and Coder 3, one of the reliability coders) agreed in more than 90% of the cases. Cohen's Kappa is also above 0.60 for all categories, indicating substantial, and in many cases even almost perfect

strength of agreement (Cohen's Kappa > .81). These results confirm high reliability of the ratings.

Table 6.8

Inter-rater reliability for coding school staff members' responses to question 3

Coder number:	1 & 2	1 & 2	1 & 2
Basic code/ sub-category	n	Percentage agreement	Cohen's Kappa
No intervention	439	99%	.80
Unspecific intervention	438	96%	.85
Intervention that exposes the new student as different	439	99%	.95
Compensatory intervention	437	98%	.75
Collect more information first	438	95%	.82
Seek advice from experts	439	97%	.78
Intervention that supports integration of the new student	439	91%	.82
Inclusive intervention	438	95%	.84
Participatory intervention	439	99%	.92

6.3. Quantitative analysis of changes in responses to the vignette between pre-test and follow-up-test

The main evaluation question that this Chapter 6 aims at answering is, whether the Hand in Hand programme had an effect on students' and school staffs' intercultural competence/diversity awareness. To answer this question, we examined changes in the frequencies of basic codes/sub-categories assigned to students' and school staff members' responses to the three questions following the description of the social scenario in the questionnaire. Students as well as school staff responded to the same scenario at T1 (pre-test) and at T3 (follow-up). The responses were coded, as described in the previous section. Codes for all categories were dichotomous: Either "yes" or "no" and one subject could receive several "yes" in multiple categories, because the coding unit was a half-sentence. Hence, to examine whether participation in the Hand in Hand programmes changed the way the participants thought about diversity, changes in the pattern of responses in different experimental groups between T1 and T3 were analysed with Chi-Square tests. Moreover, Phi-Coefficients were computed to assess the size of significant effects.

The pattern of responses of school staff was expected to change in the two experimental groups that had participated in the Hand in Hand school staff programmes (conditions C and D) and not in the other two experimental groups (condition B and control group). Accordingly, the pattern of responses of students was expected to change in the two experimental groups that had participated in the Hand in Hand student programme (experimental groups B and D), not in the control group, but possibly also in experimental group C, because school staff members and other school staff might have used their knowledge from the school staff programmes to teach intercultural competencies/diversity awareness to students.

Only participants who responded twice to the vignette were included, to avoid confounding of programme effects with changes in the samples. This meant a reduction of the total sample to 1186 (questions 1 and 2) respective 1121 (question 3) students and 318 (questions 1 and 2) respective 300 (question 3) school staff members.

6.4. Results: Changes in students' and school staff members' diversity awareness in different experimental groups

In the following, statistical results regarding differences in the patterns of responses to questions 1 and 2 (why did the new student and his classmates behave the way described in the scenario) as well as to question 3 (what would you do if you were observing this situation) are presented for students and school staff, respectively.

6.4.1. Changes in types of explanations for the behaviour of students described in the scenario suggested by students

In the following, the frequency of all categories of responses of students to questions 1 and 2 (“Why is the student behaving in that way?” and “Why are your classmates behaving in this way?”) are shown as well as their changes between T1 and T3 in different experimental groups. Chi-Square and p-values are depicted to assess whether changes are statistically significant. Moreover, Phi-Coefficients are reported to indicate effect sizes.

Table 6.9. shows that a vast majority of students (74% to 85% in each of the experimental groups at each point in time) argued that being new at the school explained why the student and his classmate behaved the way described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.9.

Number of students who argued that being new at the school was one of the reasons why the student and his classmates behaved the way described in the scenario

Being new	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	114 (85%)	107 (82%)	0.55	1	n.s.	-.05
Student programme	109 (79%)	112 (79%)	2.08	1	n.s.	.08
School staff programmes	119 (75%)	114 (74%)	0.07	1	n.s.	-.02
Student programme and school staff programmes	134 (75%)	135 (77%)	0.16	1	n.s.	.02
Total	476 (77%)	468 (78%)	0.26	1	n.s.	.01

Table 6.10. shows that only few students (1% to 7% in each of the experimental groups at each point in time) argued that the new student's violation of school norms explained why the student and his classmate behaved the way described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.10.

Number of students who argued that the new students' violation of school norms was one of the reasons why his classmates behaved the way described in the scenario

Violation school norms	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	6 (5%)	6 (5%)	0.00	1	n.s.	.00
Student programme	11 (7%)	7 (5%)	0.68	1	n.s.	-.05
School staff programmes	6 (4%)	8 (5%)	0.36	1	n.s.	.03
Student programme and school staff programmes	6 (3%)	2 (1%)	1.96	1	n.s.	-.07
Total	29 (5%)	23 (4%)	0.54	1	n.s.	-.02

Table 6.11. shows that 11% to 35% of the students in each of the experimental groups at each point in time argued that individual characteristics of the new student explained why he and/or his classmates behaved the way described in the scenario. The number of students who gave a response falling into this category became smaller between T1 and T3 in the control group and in the group that had participated in a student training only. No such change was observed in the experimental group where only the school staff had participated in the Hand in Hand programme and neither in the experimental group where students and school staff had participated in the Hand in Hand programmes. This pattern of results suggests that the significant changes were not due to participation in the programme. Moreover, the effect size (Phi-coefficient) was small.

Table 6.11.

Number of students who argued that the individual characteristics of the new student were one of the reasons why he and/or his classmates behaved the way described in the scenario

Individual characteristics of the new student	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	36 (27%)	19 (15%)	6.16	1	< .05	-.15
Student programme	33 (22%)	16 (11%)	5.76	1	< .05	-.14
School staff programmes	30 (19%)	28 (18%)	0.03	1	n.s.	-.01
Student programme and school staff programmes	31 (17%)	20 (35%)	0.42	1	n.s.	.03
Total	130 (21%)	98 (16%)	4.25	1	< .05	-.06

Table 6.12. shows that 7% to 15% of the students in each of the experimental groups at each point in time argued that characteristics of the new students' family explained why he behaved the way described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.12.

Number of students who argued that the characteristics of the new students' family were one of the reasons why he behaved the way described in the scenario

Characteristics of the new students' family	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	18 (13%)	19 (15%)	0.06	1	n.s.	.02
Student programme	10 (7%)	11 (8%)	0.15	1	n.s.	.02
School staff programmes	21 (13%)	21 (14%)	0.01	1	n.s.	.01
Student programme and school staff programmes	21 (12%)	25 (14%)	0.51	1	n.s.	.04
Total	70 (11%)	76 (13%)	0.56	1	n.s.	.02

Table 6.13. shows that 9% to 25% of the students in each of the experimental groups at each point in time argued that individual characteristics of the new students' classmates explained why they behaved the way described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.13.

Number of students who argued that individual characteristics of the new students' classmates were one of the reasons why they behaved the way described in the scenario

Individual characteristics of the classmates	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	33 (25%)	30 (23%)	0.11	1	n.s.	-.02
Student programme	23 (15%)	12 (9%)	3.12	1	n.s.	-.10
School staff programmes	28 (18%)	20 (13%)	1.34	1	n.s.	-.07
Student programme and school staff programmes	32 (18%)	28 (16%)	0.22	1	n.s.	-.03
Total	116 (19%)	90 (15%)	2.95	1	n.s.	-.05

Table 6.14. shows that 20% to 33% of the students in each of the experimental groups at each point in time explained the behaviour of the new student and/or of his classmates, described in the scenario, with emotions. The number of students who gave a response falling into this category became smaller between T1 and T3 in the experimental group where only school staff had participated in a Hand in Hand programme. As noting and recognizing the emotions of others was an important part of the Hand in Hand programme, this effect is contrary to our hypotheses. However, it also showed up only in one of the two experimental groups that had participated in a school staff programme, and the effect size (Phi-coefficient) was small.

Table 6.14.

Number of students who argued that emotions explained why the new student and his classmates behaved the way described in the scenario

Emotions	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	35 (26%)	28 (21%)	0.82	1	n.s.	-.06
Student programme	42 (28%)	28 (20%)	2.53	1	n.s.	-.09
School staff programmes	52 (33%)	33 (21%)	5.19	1	< .05	-.13
Student programme and school staff programmes	51 (29%)	45 (26%)	0.35	1	n.s.	-.03
Total	180 (29%)	134 (22%)	7.07	1	< .01	-.08

Table 6.15. shows that 11% to 20% of the students in each of the experimental groups at each point in time argued that that the new students' classmates behaved the way described in the scenario, because the new student was "strange"¹⁶. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.15.

Number of students who argued that the new students' classmates behaved the way described in the scenario, because the new student was "strange"

Strangeness of the new student	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	16 (12%)	16 (12%)	0.01	1	n.s.	.00
Student programme	30 (20%)	22 (16%)	0.91	1	n.s.	-.06
School staff programmes	27 (17%)	20 (13%)	0.98	1	n.s.	-.06
Student programme and school staff programmes	19 (11%)	21 (12%)	0.17	1	n.s.	.02
Total	92 (15%)	79 (13%)	0.67	1	n.s.	-.02

¹⁶ Please note that in most of the students' responses falling into this category it remains open, whether the new student as an individual is considered to be a "strange" person by the student or whether his status as an immigrant makes him strange in their eyes.

Table 6.16. shows that 19% to 30% of the students in each of the experimental groups at each point in time argued that an otherness of the new student explained why his classmates behaved the way described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.16.

Number of students who argued that an otherness of the new student explained why his classmates behaved the way described in the scenario

Otherness of the new student	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	25 (19%)	28 (21%)	0.31	1	n.s.	.03
Student programme	34 (23%)	42 (30%)	2.00	1	n.s.	.08
School staff programmes	37 (23%)	36 (23%)	0.00	1	n.s.	.00
Student programme and school staff programmes	40 (22%)	43 (25%)	0.24	1	n.s.	.03
Total	136 (22%)	149 (25%)	1.47	1	n.s.	.04

Table 6.17. shows that 18% to 29% of the students in each of the experimental groups at each point in time argued that a language barrier explained why the new student and/or his classmates behaved the way described in the scenario. The number of students who gave a response falling into this category became smaller between T1 and T3 in the experimental group where only school staff had participated in a Hand in Hand programme. However, this effect was not found in the second experimental group where school staff programmes had taken place (school staff programmes and student programme). Moreover, the effect size (Phi-coefficient) was small.

Table 6.17.

Number of students who argued that a language barrier explained why the new student and/or his classmates behaved the way described in the scenario

Language barrier	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	39 (29%)	33 (25%)	0.51	1	n.s.	-.04
Student programme	28 (19%)	34 (24%)	1.43	1	n.s.	.07
School staff programmes	44 (28%)	28 (18%)	3.98	1	<.05	-.11
Student programme and school staff programmes	38 (21%)	41 (23%)	0.25	1	n.s.	.03
Total	149 (24%)	136 (23%)	0.27	1	n.s.	-.02

Table 6.18. shows that 19% to 30% of the students in each of the experimental groups at each point in time argued that cultural differences explained why the new student and/or his classmates behaved the way described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.18.

Number of students who argued cultural differences explained why the new student and/or his classmates behaved the way described in the scenario

Cultural differences	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	1 (1%)	0 (0%)	0.98	1	n.s.	-.06
Student programme	4 (3%)	2 (1%)	0.55	1	n.s.	-.04
School staff programmes	7 (4%)	4 (3%)	0.77	1	n.s.	-.05
Student programme and school staff programmes	5 (3%)	4 (2%)	0.09	1	n.s.	-.02
Total	17 (3%)	10 (2%)	1.62	1	n.s.	-.04

Table 6.19. shows that 6% to 13% of the students in each of the experimental groups at each point in time argued that an ascription of otherness to the new student by his classmates explained why they behaved the way described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.19.

Number of students who argued that an ascription of otherness to the new student by his classmates explained why they behaved the way described in the scenario

Ascription of otherness	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	17 (13%)	11 (8%)	1.29	1	n.s.	-.07
Student programme	12 (8%)	12 (9%)	0.03	1	n.s.	.01
School staff programmes	12 (8%)	19 (12%)	1.96	1	n.s.	.08
Student programme and school staff programmes	11 (6%)	15 (9%)	0.76	1	n.s.	.05
Total	52 (8%)	57 (10%)	0.48	1	n.s.	.02

Table 6.20. shows that few students (0 to 1% in each of the experimental groups at each point in time) considered the role of institutional settings and practices for explaining the behaviour of the new student and/ or his classmates. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.20.

Number of students who argued that institutional settings and practices explained why the new student and/ or his classmates behaved the way described in the scenario

Institutional settings and practices	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	0	1 (1%)	1.03	1	n.s	.06
Student programme	2 (1%)	0	1.88	1	n.s	-.08
School staff programmes	0	0	-	1	n.s	-
Student programme and school staff programmes	0	0	-	1	n.s	-
Total	2 (0.3%)	1 (0.2%)	0.30	1	n.s	-.02

6.4.2. Changes in types of explanations for the behaviour of students described in the scenario suggested by school staff

In the following, the frequency of all categories of responses of members of the school staff to questions 1 and 2 (“What do you think, why is the student behaving in that way?” and “What do you think, why are his classmates behaving in this way?”) are shown as well as their changes between T1 and T3 in different experimental groups. Chi-Square and p-values are depicted to assess whether changes are statistically significant. Moreover, Phi-Coefficients are reported to indicate effect sizes.

Table 6.21. shows that between 12% and 36% of the school staff members in different experimental groups at different points in time argued that individual characteristics of the new student explained his behaviour described in the scenario. This number decreased significantly between T1 and T3 in the control group. The effect size is medium. In none of the other experimental groups a significant difference between T1 and T3 was observed.

Table 6.21.

Number of school staff members who argued that individual characteristics of the new student were one of the causes of the situation described in the scenario

Individual characteristics of the new student	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	13 (30%)	5 (12%)	4.50	1	<.05	-.23
Student programme	9 (27%)	12 (36%)	0.63	1	n.s.	.10
School staff programmes	14 (33%)	13 (30%)	0.14	1	n.s.	-.04
Student programme and school staff programmes	9 (23%)	7 (18%)	0.25	1	n.s.	-.06
Total	45 (29%)	37 (23%)	1.13	1	n.s.	-.06

Table 6.22. shows that few school staff members (0 to 9% in different experimental groups at different points in time) argued that characteristics of the new students' families explained his behaviour described in the scenario. There were further no significant differences between T1 and T3 in the number of responses assigned to this category.

Table 6.22.

Number of school staff members who argued that characteristics of the new students' family were one of the causes of the situation described in the scenario

Characteristics of the new students' families	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	4 (9%)	4 (9%)	0.00	1	n.s.	.00
Student programme	3 (9%)	0	3.14	1	n.s.	-.22
School staff programmes	3 (7%)	2 (5%)	0.27	1	n.s.	-.06
Student programme and school staff programmes	1 (3%)	1 (3%)	0.00	1	n.s.	.00
Total	11 (7%)	7 (4%)	1.00	1	n.s.	-.06

Table 6.23. shows that 18% to 47% of the school staff members in different experimental groups at different points in time argued that characteristics of the new students' classmates explained their own behaviour and/ or the behaviour of the new student described in the scenario. There were no significant differences between T1 and T3 in the number of responses assigned to this category.

Table 6.23.

Number of school staff members who argued that individual characteristics of the new students' classmates were one of the causes of the situation described in the scenario

Characteristics of the new students' classmates	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	20 (47%)	20 (47%)	0.00	1	n.s.	.00
Student programme	8 (24%)	6 (18%)	0.36	1	n.s.	-.07
School staff programmes	15 (34%)	17 (39%)	0.20	1	n.s.	.05
Student programme and school staff programmes	14 (37%)	11 (28%)	0.66	1	n.s.	-.09
Total	57 (36%)	54 (34%)	0.16	1	n.s.	-.02

Table 6.24. shows that 28% to 50% of the school staff members in different experimental groups at different points in time argued that group dynamics explained the behaviour of the new student and/ or the behaviour of his classmates described in the scenario. There were no significant differences between T1 and T3 in the number of responses falling into this category.

Table 6.24.

Number of school staff members who argued that group dynamics were one of the causes of the situation described in the scenario

Group dynamics	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	22 (51%)	17 (40%)	1.17	1	n.s.	-.12
Student programme	14 (42%)	12 (36%)	0.25	1	n.s.	-.06
School staff programmes	21 (50%)	20 (46%)	0.18	1	n.s.	-.05
Student programme and school staff programmes	16 (41%)	11 (28%)	1.42	1	n.s.	-.14
Total	73 (47%)	60 (38%)	2.49	1	n.s.	-.09

Table 6.25. shows that few school staff members (3% to 13% in different experimental groups at different points in time) argued that (post-)migration stress explained the behaviour of the new student described in the scenario. There were no significant differences between T1 and T3 in the number of responses falling into this category.

Table 6.25.

Number of school staff members who argued that (post-)migration stress was one of the causes of the situation described in the scenario

(Post-)migration stress	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	5 (12%)	4 (9%)	0.12	1	n.s.	-.04
Student programme	4 (12%)	1 (3%)	1.95	1	n.s.	-.17
School staff programmes	3 (7%)	5 (11%)	0.45	1	n.s.	.07
Student programme and school staff programmes	5 (13%)	3 (7%)	0.56	1	n.s.	-.09
Total	17 (11%)	13 (8%)	0.65	1	n.s.	-.05

Table 6.26. shows that few of the school staff members in different experimental groups at different points in time (0 to 8%) argued that an otherness of the new student explained the situation described in the scenario. There were no significant differences between T1 and T3 in the number of responses falling into this category.

Table 6.26.

Number of school staff members who argued that an otherness of the new student was one of the causes of the situation described in the scenario

Difficulties of the classmates to accept an assumed otherness	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	3 (7%)	2 (1%)	1.05	1	n.s.	-.11
Student programme	0	1 (3%)	1.02	1	n.s.	.12
School staff programmes	1 (2%)	2 (5%)	0.35	1	n.s.	.06
Student programme and school staff programmes	3 (8%)	2 (5%)	0.21	1	n.s.	-.05
Total	7 (4%)	6 (4%)	0.08	1	n.s.	-.02

Table 6.27. shows that many school staff members (40% to 72% in different experimental groups at different points in time) argued that a language barrier explained the behaviour of the new student and/ or the behaviour of his classmates described in the scenario. A significant difference between T1 and T3 in the number of responses assigned to this category was only observed for the control group. Here, a significantly smaller number of responses at T3 fell into this category as compared to T1. The effect size was medium.

Table 6.27.

Number of school staff members who argued that a language barrier was one of the causes of the situation described in the scenario

Language barrier	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	31 (72%)	17 (40%)	9.24	1	<.01	-.33
Student programme	16 (49%)	20 (61%)	0.98	1	n.s.	.23
School staff programmes	25 (57%)	23 (52%)	0.18	1	n.s.	-.05
Student programme and school staff programmes	16 (41%)	19 (49%)	0.47	1	n.s.	-.08
Total	95 (60%)	80 (50%)	2.86	1	n.s.	-.10

Table 6.28. shows that few school staff members (2% to 14% in different experimental groups at different points in time) argued that cultural differences between the new student and his classmates explained the behaviour of the new student and/ or the behaviour of his classmates described in the scenario. A significant difference between T1 and T3 in the number of responses assigned to this category was only observed for the control group. Here, a significantly smaller number of responses at T3 fell into this category as compared to T1. The effect size was medium.

Table 6.28.

Number of school staff members who argued that cultural differences between the new student and his classmates were one of the causes of the situation described in the scenario

Cultural differences	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	6 (14%)	1 (2%)	3.89	1	<.05	-.22
Student programme	3 (9%)	2 (6%)	0.22	1	n.s.	-.06
School staff programmes	5 (11%)	3 (7%)	0.55	1	n.s.	-.08
Student programme and school staff programmes	3 (8%)	3 (8%)	0	1	n.s.	.00
Total	17 (11%)	9 (6%)	2.74	1	n.s.	-.09

Table 6.29. shows that 18% to 40% of the school staff members in different experimental groups at different points in time argued that difficulties of the classmates to accept an assumed otherness of the new student explained their behaviour described in the scenario. There were no significant differences between T1 and T3 in the number of responses falling into this category.

Table 6.29.

Number of school staff members who argued that difficulties of the classmates to accept an assumed otherness of the new student was one of the causes of the situation described in the scenario

Difficulties of the classmates to accept an assumed otherness	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	10 (23%)	13 (30%)	0.53	1	n.s.	.08
Student programme	13 (40%)	8 (24%)	1.75	1	n.s.	-.16
School staff programmes	13 (30%)	8 (18%)	1.56	1	n.s.	-.13
Student programme and school staff programmes	15 (39%)	10 (26%)	1.47	1	n.s.	-.14
Total	51 (32%)	39 (25%)	2.23	1	n.s.	-.08

Table 6.30. shows that few school staff members (0% to 9% in different experimental groups at different points in time) argued that institutional settings and practices explained the new students' behaviour and/ or the behaviour of his classmates described in the scenario. In total, significantly fewer responses at T3 fell into this category. However, there was no significant difference between T1 and T3 within any of the experimental groups.

Table 6.30.

Number of school staff members who argued that institutional settings and practices were one of the causes of the situation described in the scenario

Institutional settings and practices	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	3 (8%)	2 (5%)	0.21	1	n.s.	-.05
Student programme	1 (3%)	0	1.02	1	n.s.	-.12
School staff programmes	4 (9%)	1 (2%)	1.91	1	n.s.	-.15
Student programme and school staff programmes	3 (8%)	0	3.12	1	n.s.	-.20
Total	11 (7%)	3 (2%)	4.78	1	<.05	-.12

6.4.3. Changes in types of interventions suggested by students to solve the situation described in the scenario

In the following, the frequency of all categories of responses of students to question 3 (“What would you do if you were observing this situation in your school?”) are shown as well as their changes between T1 and T3 in different experimental groups. Chi-Square and *p*-values are depicted to assess whether changes are statistically significant. Moreover, Phi-Coefficients are reported to indicate effect sizes.

Table 6.30. shows that 5% to 17% of students in each of the experimental groups at each point in time said that they would do nothing if they were observing the situation described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.30.

Number of students who said they would not intervene if they were observing the situation described in the scenario

No intervention	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	9 (7%)	9 (7%)	0.00	1	n.s.	.00
Student programme	7 (5%)	9 (7%)	0.35	1	n.s.	.04
School staff programmes	27 (17%)	17 (11%)	2.49	1	n.s.	-.09
Student programme and school staff programmes	18 (10%)	23 (13%)	0.83	1	n.s.	.05
Total	61 (10%)	58 (10%)	0.03	1	n.s.	.00

Table 6.31. shows that few students (1% to 5% in each of the experimental groups at each point in time) said that they do not know what they would do if they were observing the situation described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.31.

Number of students who said they do not know how they would react if they were observing the situation described in the scenario

Unsure	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	6 (5%)	9 (7%)	0.64	1	n.s.	.05
Student programme	7 (5%)	9 (7%)	0.35	1	n.s.	.04
School staff programmes	5 (3%)	8 (5%)	0.76	1	n.s.	.05
Student programme and school staff programmes	7 (4%)	6 (4%)	0.06	1	n.s.	-.01
Total	25 (4%)	32 (5%)	1.04	1	n.s.	.03

Table 6.32. shows that few students (1% to 5% in each of the experimental groups at each point in time) said that they would be friendly but would not intervene if

they were observing the situation described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.32.

Number of students who said they would be friendly but would not intervene if they were observing the situation described in the scenario

Friendly, but no intervention	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	1 (1%)	2 (2%)	0.34	1	n.s.	.04
Student programme	4 (3%)	3 (2%)	0.11	1	n.s.	-.02
School staff programmes	8 (5%)	6 (4%)	0.27	1	n.s.	-.03
Student programme and school staff programmes	3 (2%)	4 (2%)	0.17	1	n.s.	.02
Total	16 (3%)	15 (3%)	0.02	1	n.s.	.00

Table 6.33. shows that few students (3% to 5% in each of the experimental groups at each point in time) said that they would ask an adult (mostly teacher, sometimes also school psychologist or parents) for help if they were observing the situation described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.33.

Number of students who said they would ask an adult for help if they were observing the situation described in the scenario

Ask adult for help	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	6 (5%)	7 (5%)	0.08	1	n.s.	.02
Student programme	4 (3%)	7 (5%)	0.97	1	n.s.	.06
School staff programmes	5 (3%)	5 (3%)	.00	1	n.s.	.00
Student programme and school staff programmes	9 (5%)	7 (4%)	0.22	1	n.s.	-.03
Total	24 (4%)	26 (4%)	0.13	1	n.s.	.01

Table 6.34. shows that 25% to 39% of the students in each of the experimental groups at each point in time said that they would help the new student if they were observing the situation described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.34.

Number of students who said they would help the new student if they were observing the situation described in the scenario

Help the new student	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	51 (39%)	37 (28%)	3.34	1	n.s.	-.11
Student programme	45 (32%)	45 (33%)	0.04	1	n.s.	.01
School staff programmes	44 (28%)	45 (29%)	0.04	1	n.s.	.01
Student programme and school staff programmes	45 (25%)	49 (28%)	0.38	1	n.s.	.03
Total	185 (30%)	176 (29%)	0.13	1	n.s.	-.01

Table 6.35. shows that many of the students (50% to 67% in each of the experimental groups at each point in time) said that they would try to connect with the new student if they were observing the situation described in the scenario. A significant difference between both points in time (T1 and T3) was observed in

the experimental group that had participated in school staff programmes. In this experimental group the percentage of students who said that they would try to connect with the new student in the situation described increased from 50% to 62%. However, no such difference was observed in the second experimental group that had participated in school staff programmes (as well as the student programme). Neither was such a difference observed in the experimental group that had participated in a student programme only. Hence, there is no evidence that the observed change between both time points is caused by the Hand in Hand programme.

Table 6.35.

Number of students who said they would try to connect with the new student if they were observing the situation described in the scenario

Connect with the new student	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	82 (62%)	82 (62%)	0.00	1	n.s.	.00
Student programme	96 (67%)	89 (65%)	0.22	1	n.s.	-.03
School staff programmes	80 (50%)	98 (62%)	4.71	1	< .05	.12
Student programme and school staff programmes	101 (57%)	105 (61%)	0.48	1	n.s.	.04
Total	359 (59%)	374 (62%)	1.62	1	n.s.	.04

Table 6.36. shows that between 10% and 22% of the students in each of the experimental groups at each point in time said that they would address group dynamics in the class if they were observing the situation described in the scenario. This percentage decreased in the experimental group that had participated in a student programme as well as in school staff programmes. This effect is in contrast to our expectations, because the Hand in Hand programmes are assumed to increase the awareness of group dynamics and/or students' self-efficacy for influencing group dynamics in a positive way. However, neither in the experimental group that had participated in a student programme only, nor in the experimental group that had participated in school staff programmes only a similar unexpected effect was observed. Moreover, the effect size was small.

Table 6.36.

Number of students who said they would address group dynamics in the class if they were observing the situation described in the scenario

Address group dynamics in the class	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	29 (22%)	27 (21%)	0.09	1	n.s.	-.02
Student programme	19 (13%)	19 (14%)	0.01	1	n.s.	.01
School staff programmes	29 (18%)	25 (16%)	0.30	1	n.s.	-.03
Student programme and school staff programmes	37 (21%)	17 (10%)	8.23	1	<.01	-.15
Total	114 (19%)	88 (15%)	3.47	1	n.s.	-.05

Table 6.37. shows that between 9% and 14% of the students in each of the experimental groups at each point in time said that they would ask the new student about his perspective if they were observing the situation described in the scenario. There were no significant differences between T1 and T3 in any of the experimental groups.

Table 6.37

Number of students who said they would try to ask the new student about his perspective if they were observing the situation described in the scenario

Ask the new student about his perspective	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	13 (10%)	18 (14%)	0.91	1	n.s.	.06
Student programme	19 (13%)	12 (9%)	1.51	1	n.s.	-.07
School staff programmes	22 (14%)	22 (14%)	1.00	1	n.s.	.00
Student programme and school staff programmes	20 (11%)	24 (14%)	0.53	1	n.s.	.04
Total	74 (12%)	76 (13%)	.09	1	n.s.	.01

6.4.4. Changes in types of interventions suggested by school staff to solve the situation described in the scenario

In the following, the frequency of all categories of responses of members of the school staff to question 3 (“What would you do if you were observing this situation in your school?”) are shown as well as their changes between T1 and T3 in different experimental groups. Chi-Square and p-values are depicted to assess whether changes are statistically significant. Moreover, Phi-Coefficients are reported to indicate effect sizes.

Table 6.38. shows that 8% to 29% of participants in each of the experimental groups at each point in time gave an unspecific answer to question 3 following the scenario. That is, they said they would do something if they were observing the situation described in the scenario, but they did not specify what they would do. Differences between T1 and T3 in the number of participants who gave unspecific answers to question 3 were not significant in any of the experimental groups.

Table 6.38.

Number of school staff members who said they would do something, but did not specify what they would do if they were observing the situation described in the scenario

Unspecific intervention	T1	T3	χ^2	df	p	ϕ
Control group	7 (16%)	11 (26%)	1.12	1	n.s.	.11
Student programme	6 (21%)	7 (24%)	0.10	1	n.s.	.04
School staff programmes	3 (8%)	7 (18%)	1.83	1	n.s.	.15
Student programme and school staff programmes	8 (21%)	11 (29%)	0.63	1	n.s.	.09
Total	24 (16%)	36 (24%)	3.00	1	n.s.	.10

Table 6.39. shows that few members of the school staff (0 to 15% in each of the experimental group at each point in time) suggested interventions that would expose the new student and address him as the (cultural) “other” (such as suggesting that he should present himself and his heritage culture in the classroom). There are no significant differences between T1 and T3 with regard

to the frequency with which school staff suggests an intervention falling into this category in any of the experimental groups.

Table 6.39.

Number of school staff members who suggested an intervention that would expose the new student and address him as the (cultural) “other”

Intervention that exposes the new student and addresses him as the (cultural) “other”	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	5 (12%)	2 (5%)	1.40	1	n.s.	-.13
Student programme	1 (3%)	0	1.02	1	n.s.	-.13
School staff programmes	6 (15%)	2 (5%)	2.22	1	n.s.	-.17
Student programme and school staff programmes	3 (8%)	3 (8%)	0.00	1	n.s.	.00
Total	15 (10%)	7 (5%)	3.14	1	n.s.	-.10

Table 6.40. shows that only few members of the school staff (9 to 7% in each of the experimental groups at each point in time) suggested compensatory interventions (aimed at addressing an ascribed deficit of the new student, such as a lack of skills in the language of instruction or a lack of knowledge of the school rules, etc.). No difference between T1 and T3 was observed in any of the experimental groups.

Table 6.40.

Number of school staff members who suggested a compensatory intervention

Compensatory intervention	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	4 (9%)	2 (5%)	0.72	1	n.s.	-.09
Student programme	0	2 (7%)	2.07	1	n.s.	.19
School staff programmes	0	3 (4%)	3.12	1	n.s.	.20
Student programme and school staff programmes	0	0	-	-	-	-
Total	4 (3%)	7 (5%)	0.85	1	n.s.	.05

Table 6.41. shows that between 3% and 20% of school staff members in each of the experimental groups at each point in time said that they would try to collect more information first if they were observing the situation described in the scenario. Differences between T1 and T3 in the number of participants who gave a response falling in this category were not significant in any of the experimental groups.

Table 6.41

Number of school staff members who said they would collect more information first if they were observing the situation described in the scenario

Collect more information	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	7 (16%)	5 (12%)	0.39	1	n.s.	-.07
Student programme	5 (17%)	5 (17%)	0.00	1	n.s.	.00
School staff programmes	8 (20%)	6 (15%)	0.35	1	n.s.	-.07
Student programme and school staff programmes	1 (3%)	4 (11%)	1.93	1	n.s.	.16
Total	21 (14%)	20 (13%)	0.28	1	n.s.	-.01

Table 6.42. shows that 3% to 12% of the participants within each experimental group at each point in time would seek advice from experts (e.g. members of the student health team, pedagogues, social workers, school psychologist, external advice centres), if they were observing the situation described in the scenario. No significant changes between T1 and T3 in the number of responses falling into this category were observed.

Table 6.42.

Number of school staff members who said they would seek advice from experts if they were observing the situation described in the scenario

Seek advice from experts	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	2 (5%)	5 (12%)	1.40	1	n.s.	.13
Student programme	3 (10%)	1 (3%)	1.07	1	n.s.	-.14
School staff programmes	1 (3%)	3 (8%)	1.05	1	n.s.	.12
Student programme and school staff programmes	4 (11%)	1 (3%)	1.93	1	n.s.	-.16
Total	10 (7%)	10 (7%)	0.00	1	n.s.	.00

Table 6.43. shows that – across experimental groups – a majority of members of the school staff suggested an intervention that aimed at supporting the integration of the new student into the class, but was not inclusive. More specifically, they suggested measures that either addressed both sides (the new student and his classmates) or only the classmates of the new student, and that had the aim to increase the new students' acceptance by the group, yet reflect a two-group theory (majority group vs. student in need) and expose the new student as someone with special needs – which is characteristic of integrative approaches to address diversity. At T3 significantly less members of the school staff gave responses that fell into this category. However, a significant decrease is only significant in the control group. We would have expected to observe such a decrease and, at the same time, an increase in inclusive suggestions in the two experimental groups that had participated in school staff trainings, but not in the control group. Hence, there is no evidence that the Hand in Hand intervention caused this change.

Table 6.43.

Number of school staff members who suggested an intervention that supports integration of the new student

Intervention that supports integration of the new student	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	23 (55%)	13 (31%)	4.86	1	< .05	-.24
Student programme	18 (62%)	12 (41%)	2.49	1	n.s.	-.21
School staff programmes	22 (55%)	19 (48%)	0.45	1	n.s.	-.08
Student programme and school staff programmes	21 (55%)	16 (42%)	1.32	1	n.s.	-.13
Total	84 (56%)	60 (40%)	7.74	1	<.01	-.16

Table 6.44. shows that between 10% and 29% of the school staff members in each experimental group at each point in time suggested an inclusive intervention, i.e. an intervention that addressed the whole class without exposing the new student as being different in any way (e.g. using cooperative learning or social games). In total significantly more school staff members suggested interventions falling into this category at T3 as compared to T1. However, the largest increase was observed in the control group and within each experimental group, including the control group, the change was not significant. Hence, there is no evidence that changes in the frequency of this type of response are due to participation in the Hand in Hand programme.

Table 6.44.

Number of school staff members who suggested an inclusive intervention

Inclusive intervention	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	5 (12%)	12 (29%)	3.61	1	n.s.	.21
Student programme	3 (10%)	6 (21%)	1.18	1	n.s.	.14
School staff programmes	6 (15%)	11 (28%)	1.87	1	n.s.	.15
Student programme and school staff programmes	4 (11%)	4 (11%)	0.00	1	n.s.	.00
Total	18 (12%)	33 (22%)	5.32	1	<.05	.13

Table 6.45 shows that only few members of the school staff (0 to 5% in each of the experimental groups at each point in time) suggested a participatory intervention (i.e., they said that they would involve the students themselves in finding a way to improve the situation described in the scenario). There was no significant difference between T1 and T3 in the number of responses falling into this category.

Table 6.45.

Number of school staff members who suggested a participatory intervention

Participatory intervention	T1	T3	χ^2	df	<i>p</i>	ϕ
Control group	0	0	-	-	-	-
Student programme	1 (3%)	0	1.02	1	n.s.	-.13
School staff programmes	1 (3%)	0	1.01	1	n.s.	-.11
Student programme and school staff programmes	2 (5%)	1 (3%)	0.35	1	n.s.	-.07
Total	4 (3%)	1 (1%)	1.83	1	n.s.	-.08

6.5. Conclusions

This Chapter 6 aimed at examining effects of the Hand in Hand programmes on students' and school staffs' intercultural competence/ diversity awareness. We compared changes in participants' interpretations of a social situation, which triggered beliefs about the social difference category "migration background" with those of non-participants, as well as changes in participants' repertoire of strategies for dealing with the situation between two measurement points with that of non-participants. We expected to find changes in the response patterns of school staff in those two experimental groups that had participated in the Hand in Hand school staff programmes (Conditions C and D), but not in the experimental group that had participated in a student programme only (Condition B) and neither in the control group (Condition A). We further expected to find changes in the response patterns of students in those experimental groups that had participated in the Hand in Hand student programme (Conditions B and D) and possibly also in the experimental group that participated in the school staff programmes only (Condition C), but not in the control group (Condition A). In particular, we expected that participants would consider cultural differences as reasons for the behaviour of students described in the scenario more often after participating in the Hand in Hand programme, because the programme included exercises that aimed at supporting reflection of cultural influences on behaviour increase in intercultural competence). And we expected, that participants would consider relations of inequality and exclusion mechanisms as well as institutional barriers as reasons for the behaviour of students described in the scenario more often after participating in the Hand in Hand programme, because the programme included exercises that aimed at supporting reflection of societal power relations. Additionally, we expected that students who had participated in a Hand in Hand programme would suggest more effective solutions for the situation after the programme as compared to before, while we expected no such changes or smaller changes in the control group. Finally, we expected that school staff who had participated in a Hand in Hand programme would suggest more inclusive solutions for the situation after the programme as compared to before, while we expected no such changes or smaller changes in the control group.

The expected pattern of changes in different experimental groups was not observed for any of the basic codes/sub-categories. Generally, students as well as school staff responded in a largely similar way to the questions at both time points independent of participation in a Hand in Hand programme. Across experimental groups, we observed few significant changes in the frequencies for each of the basic codes/ sub-categories. Several of the few significant differences between T1

and T3 we had observed were even in opposition to our expectations. None of the effects was found in more than one experimental group. Hence, the vignette study provides no evidence for an effect of participation in the Hand in Hand programmes on students' or school staffs' intercultural competence/ diversity awareness. This is in accordance with the self-report-based evaluation results presented in Chapter 5: None of the methods used in a mixed-methods study created evidence of an effect of the Hand in Hand programmes on students' or school staffs' intercultural competencies/ diversity awareness.

6.6. References

Barter, C., & Renold, E. (2000). "I Wanna Tell You a Story" Exploring the Application of Vignettes in Qualitative Research with Children and Young People. *International Journal of Social Research Methodology*, 3, 307-323.

Busse, V., & Krause, U.-M. (2015). Addressing Cultural Diversity: Effects of a problem-based intercultural learning unit. *Learning Environments Research*, 18(3), 425 – 452.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105, 98-121.

He, J., & Van de Vijver, F. J. R. (2013). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences*, 55, 794-800.

Hesse, H. G. & Göbel, K. (2007). Interkulturelle Kompetenz. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen – Konzepte und Messung – DESI-Studie* (S. 253-269). Weinheim: Beltz.

Hill, M. (1997). Research Review: Participatory Research with Children. *Child and Family Social Work*, 2, 171–183.

Hughes, R. (1998). Considering the vignette technique and its application to a study of drug injecting and HIV risk and safer behaviour. *Sociology of Health and Illness*, 20, 381-400

Jackson, D. N., Messick, S. J. (1965). Acquiescence: The nonvanishing variance component. *American Psychologist*, 20, 498.

Keefer, K. V. (2014). Self-Report Assessments of Emotional Competencies: A Critical Look at Methods and Meanings. [Editorial]. *Journal of Psychoeducational Assessment*, 33(1), 3–23.

Knowles, E. S., Condon, C. A. (1999). Why people say “yes”: A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, 77, 379-386.

Krosnick, J. A., Presser, S. (2010). Question and questionnaire design. In Marsden, P. V., Wright, J. D. (Eds.), *Handbook of survey research* (2nd ed., pp. 263-314). Bingley, UK: Emerald.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York: The Guilford Press.

Richman, J. and Mercer, D. (2002). The vignette revisited: Evil and the forensic nurse. *Nurse Researcher*, 9 (4), 70–82.

Rockstuhl, T., Ang, S., Ng, K. Y., Lievens, F., Van Dyne, L. (2015). Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *Journal of Applied Psychology*, 100, 464-480.

Schoenberg, N. E., & Ravdal, H. (2000). Using vignettes in awareness and attitudinal research. *International Journal of Social Research Methodology*, 3, 63-74.

Schwarzenthal, M., Juang, L., Schachner, M. K., & van de Vijver, F. J. R. (2019). A multimodal measure of cultural intelligence for adolescents growing up in culturally diverse societies. *International Journal of Intercultural Relations*, 72, 109-12.

Syed, M., & Nelson, S. C. (2015). Guidelines for establishing reliability when coding narrative data. *Emerging adulthood*, 3(6), 375-387.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19(3), 188–202.

Chapter 7:
**Effects of the HAND in HAND programmes on classroom and
school climates**

Svenja Vieluf, Mojca Rožman, Nina Roczen

7.1. Introduction

The classroom climate refers to students' and/or teachers' shared perception of the quality of the classroom environment (Adelmann & Taylor, 2005; Cohen, McCabe, Michelli, & Pickeral, 2009; Moos, 1973; van Houtte, 2005; Walberg & Anderson, 1968), in particular, of the quality of interactions among students and between students and teachers in the classroom (see e.g. Moos, 1973). The school climate has been defined as “the quality and character of school life [that] is based on patterns of people’s experiences of school life and reflects norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures” (Cohen et al. 2009, p. 182). Since relationship skills, i.e., “the ability to establish and maintain healthy and rewarding relationships with diverse individuals and groups“, are an important part of social competency (CASEL, 2013), it can be expected that a programme addressing students' and teachers' social competencies will also positively affect the quality of relations in school and, thus, the classroom and school climates (see also Chapter 1).

In this chapter, the hypothesis that the HAND in HAND programmes had a positive effect on classroom and school climates, is examined. Schools/classes¹⁷ from the different experimental groups (control group, student programme only, staff programmes only, student and staff programmes) are compared with regard to changes in their school/classroom climates between the measurement points T1 (before the programme), T2 (after the programme), and T3 (follow-up).

To measure classroom climates, the HAND in HAND evaluation used a multi-method approach. This had the aim to increase the validity of conclusions. The following types of instruments were used: questionnaire scales, a sociometric instrument, and semi-structured focus-group interviews with participants.

- **Questionnaire scales:** At each of the three measurement points several questionnaire scales were used to assess students' and teachers' perception of the classroom climate. The questionnaire scales asked about three sub-dimensions of classroom climates: the quality of student-student-relations, the quality of student-teacher relations, and the disciplinary climate in the classroom.
- **Sociometric instrument:** At each of the three measurement points students were additionally asked to list the names of those students with whom they had most commonly spent their breaks/recess with during the 4 months

¹⁷ Because only one class per school participated in the training, both levels coincide. In Sweden the dropout between T2 and T3 was too large so that no data for Sweden for T3 was available.

preceding the respective assessment and to list the names of those students with whom they did not spend any of their breaks/recess during the 4 months preceding the respective assessment. Based on the responses to these two questions three classroom-level network-indicators were computed (based on Hennig, Brandes, Pfeffer, & Mergel, 2012 and Jansen, 2006): (a) the density of the social network in the classroom (number of reported relations divided by the number of possible relations in a class;); (b) the percent of isolated students (students who did not report to spend their breaks commonly with any other students from the class) and (c) the percent of unpopular students (students that were named by no other students or by only one other student in response to the question with whom they commonly spent their breaks) were investigated.

- ***Semi-structured focus group interviews:*** During semi-structured focus-group interviews that took place after the programme had been completed (see Chapter 2 and see also Chapter 12 and Chapter 12), students and teachers were asked whether they had noticed any changes to the climate of their class/the target class in the respective school.

For theoretical reasons and because different types of indicators were derived from different data sources, changes in classroom climates were examined at two different levels: the student and the classroom level. In previous research classroom climates have often been conceptualized as the “shared perception” (Moos, 1973) of students and/or teachers and, hence, as a classroom level construct. Yet, the qualities of relationships between individuals within the system of the classroom can vary considerably (e.g. Davis, 2003; Rosenthal, 1994). Further, it is possible that the HAND in HAND programme helped only some individuals and not others to improve their social competencies, which would imply that programme effects could be expected for specific relations within the classroom only and not necessarily for the whole system. Therefore, students’ reports of the quality of their relationships with teachers and with other students can be expected to vary at both, the student and the classroom levels. To take this into account, student responses to questionnaire questions concerning the quality of their individual relations with teachers and classmates were analyzed at the individual level and separately for the three language versions of the questionnaire. In contrast, teachers’ reports of student-student relations were analyzed at the classroom level. Also, questions in the student questionnaire that concerned classroom processes (discrimination of minority students and the disciplinary climate) were analyzed at the classroom level for theoretical reasons. It should be noted that the sample at the classroom level was small (max. 36

classes/schools altogether and only 12 classes/schools per language version of the questionnaire and for some measures even less). Therefore, analysis was done across language versions of the questionnaire when it focused on the classroom/school level.

The analysis of student-level effects followed the procedure lined out in chapter 5. To examine effects of the HAND in HAND programme at the classroom-level, variance analysis for repeated measurements was carried out with the programme SPSS. Main effects of the measurement point and of the experimental condition are reported as well as interaction effects between both. In contrast, interview responses about changes in classroom climates since the beginning of the school year were analysed with qualitative content analysis. The inductive categories were summarized, counted and illustrated with quotes from the interviews. First-order codes were additionally enumerated and Chi-Square tests were used to examine differences between experimental groups with regard to the existence of observed changes in classroom climates.

Changes to the school climate were not so much in the focus of the external evaluation of the HAND in HAND programme (because only one class per school and only a small group of school staff could participate in the trainings and, thus, it could not be expected that the whole school would change considerably). Yet, one question during the semi-structured focus-group interviews, also asked about one aspect of the school climate: changes in the quality of relations among school staff. The analysis procedure was the same as for the semi-structured focus-group interview questions concerning the classroom climate.

The present chapter is structured as follows: The first section concerns changes in the quality of students-student relations, the second section addresses changes in the quality of student-teacher relations, and the third section changes in disciplinary classroom climates. The fourth section is about the social structure of the classroom and discusses sociometric results. The fifth section summarizes responses of participants during semi-structured focus-group interviews about their perception of changes in classroom climates and in school climates, i.e. the quality of relations among staff in different experimental groups. The chapter ends with a summary and integration of results and with a conclusion.

7.2. Changes in the quality of student-student relations

Information on changes in the quality of student-student relations comes from the student and the teacher questionnaires. Students were asked how often they experienced bullying by their classmates during the 4 months preceding the

respective assessment. Teachers were asked whether they had observed verbal and/or physical violence among students in the target class and how they would assess the quality of student-student relations in the target class on a more general level. Student and teacher responses are presented separately in the following.

7.2.1. Student Reports

Student reports of bullying were analysed at the student level. Figure 7.1. shows differences between T2 and T1 scale scores for bullying for all four experimental groups. Figure 7.2. shows the same for differences between T1 and T3.

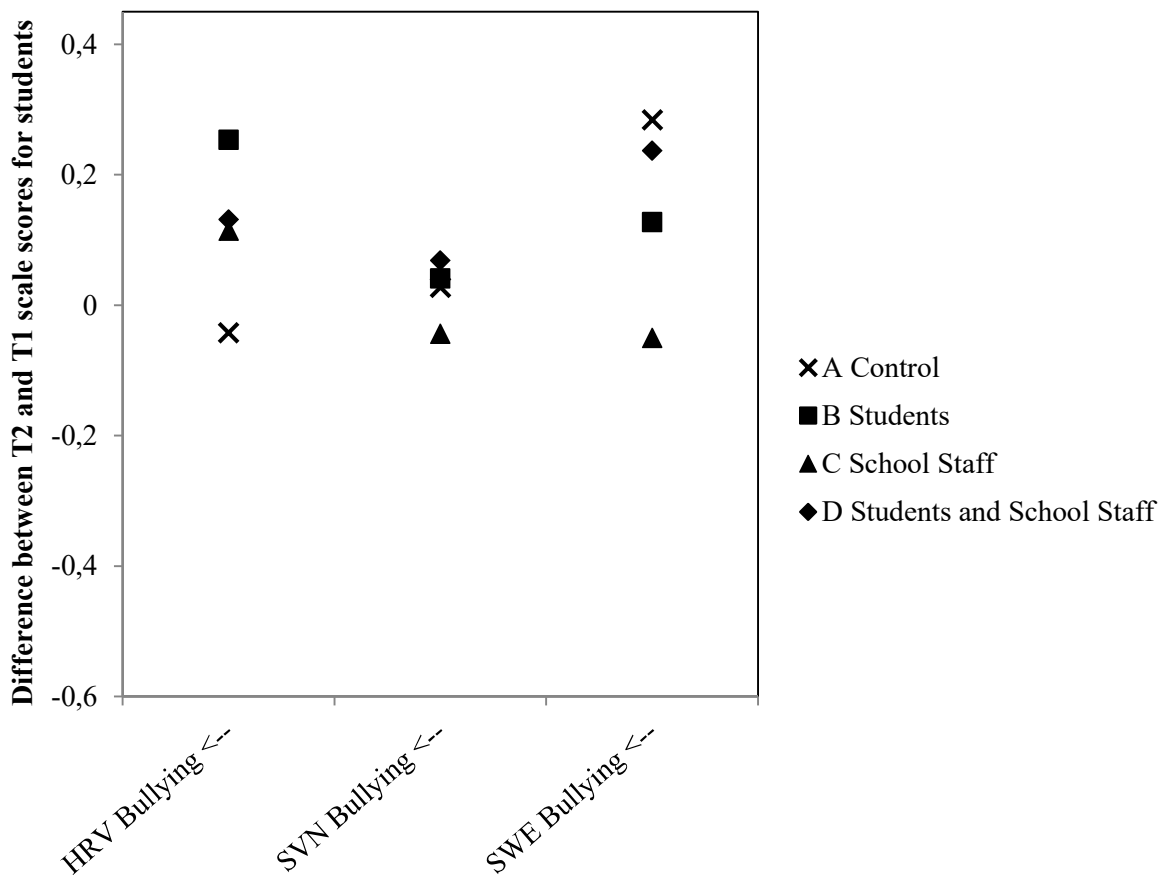


Figure 7.1. Average difference between T2 and T1 per group for student reports of the quality of student-student relations (perceived frequency of bullying) in Croatia, Slovenia, and Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

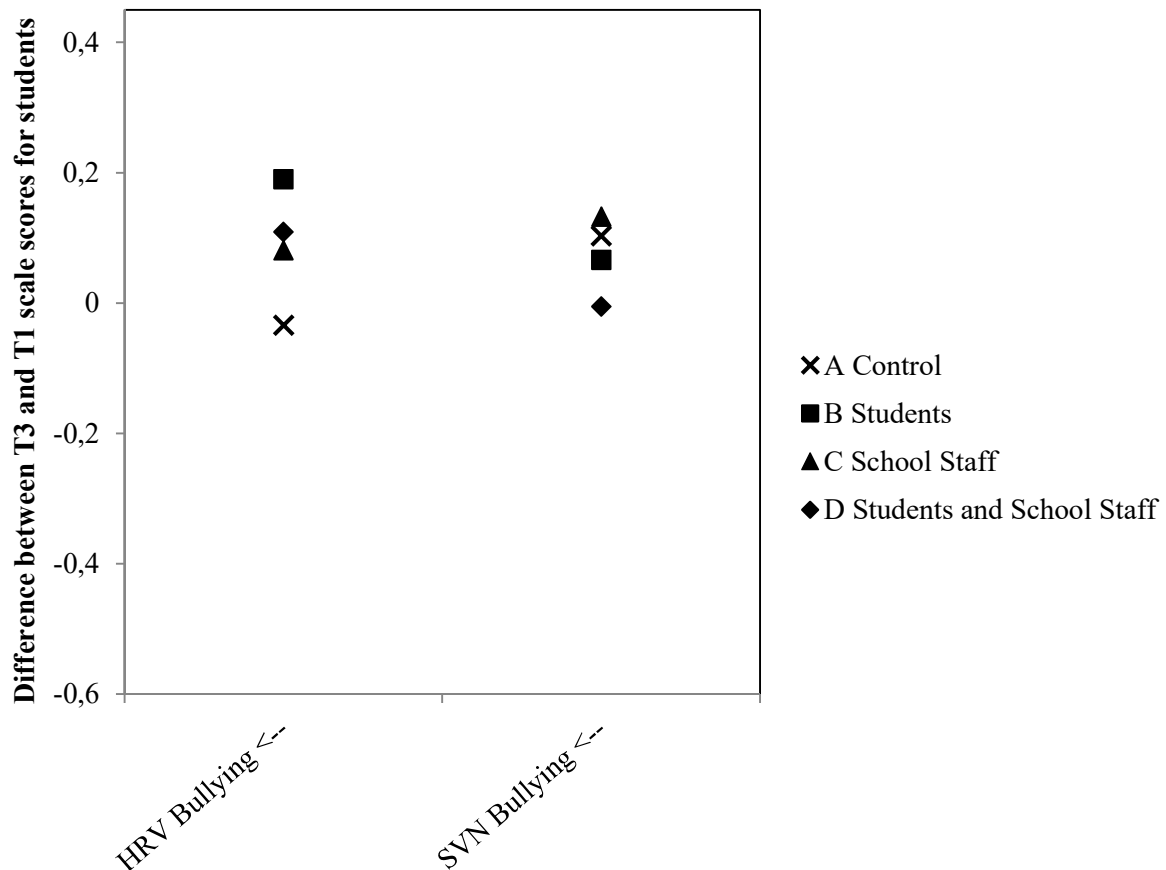


Figure 7.2. Average difference between T3 and T1 per group for student reports of the quality of student-student relations (perceived frequency of bullying) in Croatia, Slovenia, and Sweden

Note: the arrow following the scale name on the x-axis points in the direction of the expected effect.

Figures 6.1. and 6.2. show that effects of the experimental group on the manifest difference score for changes in the frequency of bullying between measurement points T1 and T2 differed considerably between school systems:

- In Croatia a significant unexpected effect of the treatment on the manifest difference between T1 and T2 concerning student-perceived bullying is observed: The frequency of being bullied increased between T1 and T2 and this increase was significantly larger in experimental groups B and C (i.e. in schools where students had participated in a HAND in HAND student programme and in schools where the school staff had participated in a school staff training) as compared to the control group (for group B: $t=2.45$, $p=0.015$, $d=0.39$; for group C: $t=2.67$, $p=0.008$, $d=0.21$). Yet, the effect sizes were small.
- In Slovenia, one of the four expected effect of the treatment on the manifest difference between T1 and T2 concerning student-perceived bullying was significant and in the expected direction: The average difference between

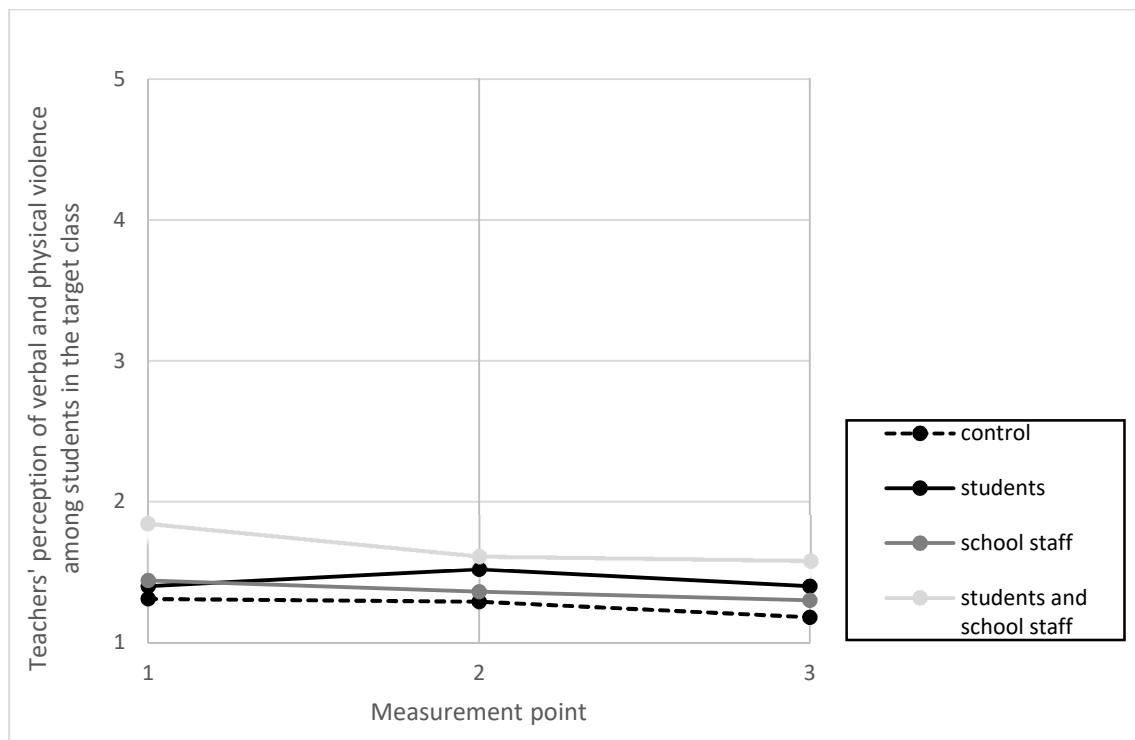
the two points in time is bigger for group C than for the control group ($t=-2.02$, $p=0.044$, $d=-0.11$). While the scale score increased in the control group and groups B and D (which implies more bullying), it decreased in group C. However, this effect is very small.

- In Sweden no significant effect of the experimental group on the difference between T1 and T2 regarding the frequency of bullying was found.¹⁸

In summary, there are single significant effects of the HAND in HAND programmes on the manifest change in bullying between T1 and T2 in single school systems that go in opposite directions. Findings concerning longer-term changes in bullying in different experimental groups are more consistent: In none of the school systems was an effect of the experimental group on the difference between T1 and T3 observed. Hence, there is no evidence of a long-term effect of the HAND in HAND programmes on the frequency of bullying in the classroom.

7.2.2. Teacher Reports

The reports of different teachers teaching the same class about the frequency of verbal and physical violence among students of the target class and about the quality of student-student relations in that class more generally were aggregated and analysed at the classroom level. The results are shown in figures 6.3. and 6.4.



No results for T3 were available for Sweden.

Figure 7.3. Differences between experimental groups in teachers' perception of verbal and physical violence among students in the target class

Figure 7.3. shows that verbal or physical violence among students as perceived by the teachers was seldom: the questionnaire scale had a minimum at “1 – not at all” and a maximum at “5 – several times daily” and the means for all experimental groups were below “2”. There were, however, significant differences between experimental groups ($F=3.80$, $df=3$, $p=.02$, $partielles \eta^2=.29$). The effect size even suggests a strong effect. Posthoc tests show that teachers teaching classes that participated in both, the HAND in HAND student programme and the HAND in HAND school staff programme reported significantly more verbal and/or physical violence among students than teachers teaching classes that belonged to the control group. This did not change after the programme: Neither had the time a significant main effect ($F=1.99$, $df=1$, $p=0.15$), nor was the interaction between time and experimental group significant ($F=0.71$, $df=3$, $p=0.64$).

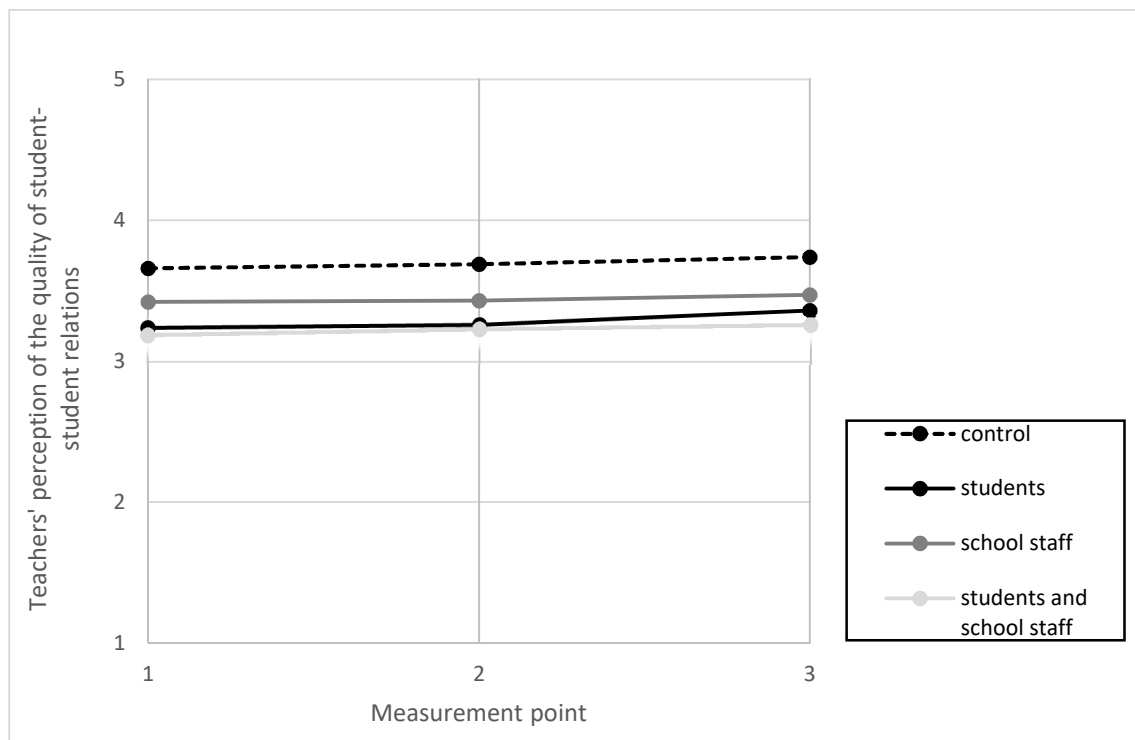


Figure 7.4. Differences between experimental groups in teachers' reports of the quality of student-student relations in the target class

Figure 7.4. shows that teachers assess the quality of student-student relations in the target classes as good. There was no significant difference between experimental groups ($F=2.46$, $df=3$, $p=.08$) and no significant change over time ($F=3.03$, $df=1$, $p=0.06$). Most importantly, the development of student-student relations over time was largely parallel in the four experimental groups: the

interaction between time and experimental group was also not significant ($F=0.14$, $df=3$, $p=0.99$).

In summary, there is no evidence of long-term effects of participation in the HAND in HAND student and/or school staff programmes on the quality of student- student-relations – irrespective of whether the analysis are based on student or teacher reports. There might be short term effects of the trainings, but these are rather difficult to interpret. In Croatia an unexpected negative effect of participation in the student training was found as well as an unexpected negative effect of participation in the school staff trainings. In contrast, participation in both trainings had no effect. In Slovenia participation in the school staff training had a positive effect, but participation in the school staff and the student training had no effect. It remains open, whether these results can actually be attributed to programme participation or whether other factors might have caused these differences between experimental groups. In any case, the observed effects of the experimental group on differences between T1 and T2 do not appear to be sustainable – until the follow-up measurement they had vanished.

7.3. Changes in the quality of student-teacher relations

The student questionnaire contained two questions asking about students' perception of the quality of their individual relations with the teachers. The scale perceived quality of student-teacher relations (positively worded) asked students whether they had experienced different forms of recognition from their teachers. The scale perceived quality of student-teacher relations (negatively worded) asked students whether they had experienced different forms of misrecognition from their teachers. Responses to these two questions were analysed at the individual student level, because variance between students in the same class is likely to reflect – besides error variance that can never be precluded – also real differences in the quality of individual relations between students and teachers (teacher usually get along better with some students than with others). A third question in the student questionnaire was analysed at the classroom level, because this question asked about students' shared perception of teacher behaviour. That is, students were asked whether they think that their teachers' discriminate on the basis of ascribed ethnicity or culture. Here, differences in responses between students assessing the same teachers might reflect differences in what students have observed, differences in students' sensitivity to discriminatory behaviour, or error variance, but not variation in individual student-teacher relationship quality. Therefore, students' responses to this scale are aggregated to the class level and differences between experimental groups are analysed at this level.

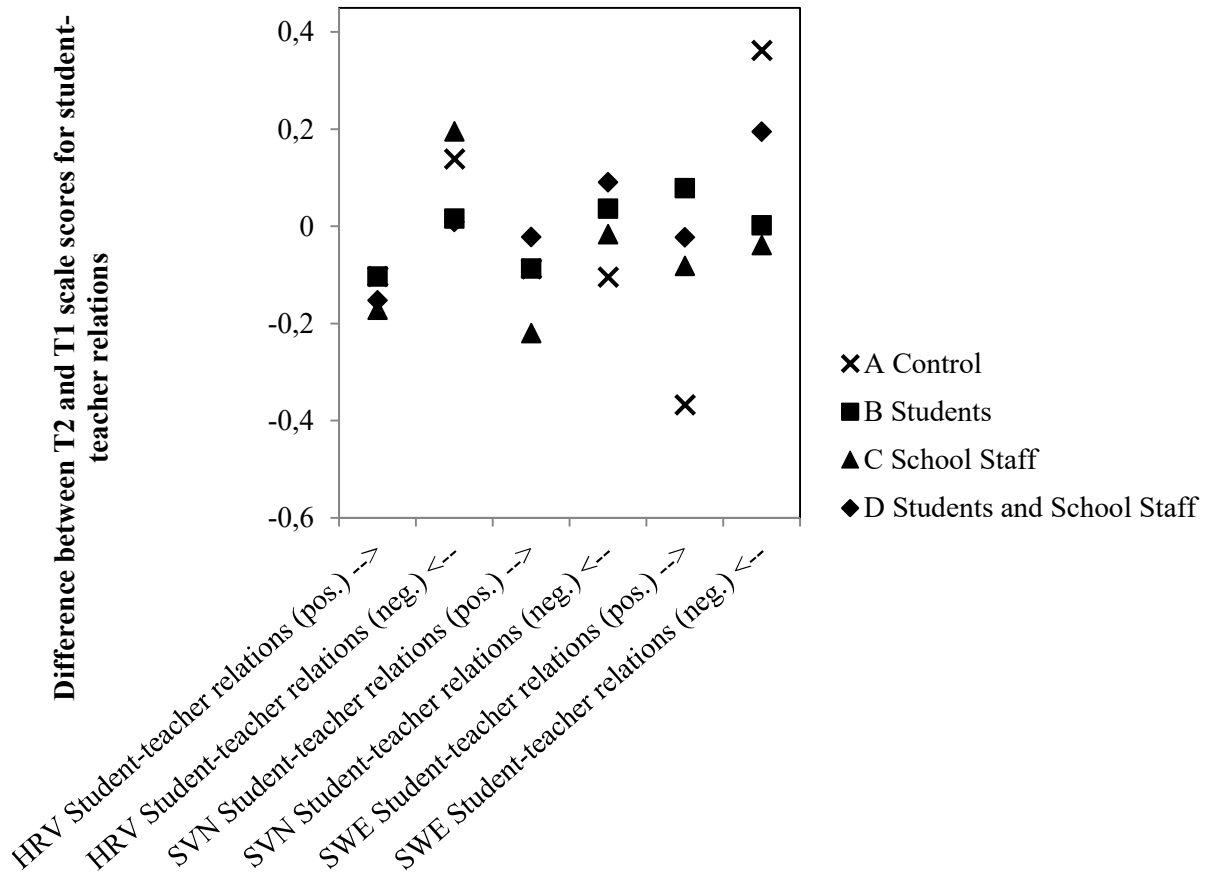


Figure 7.5. Average difference between T2 and T1 per group for student reports of the quality of student-teacher relations (recognition and misrecognition of students by their teachers) in Croatia, Slovenia, and Sweden

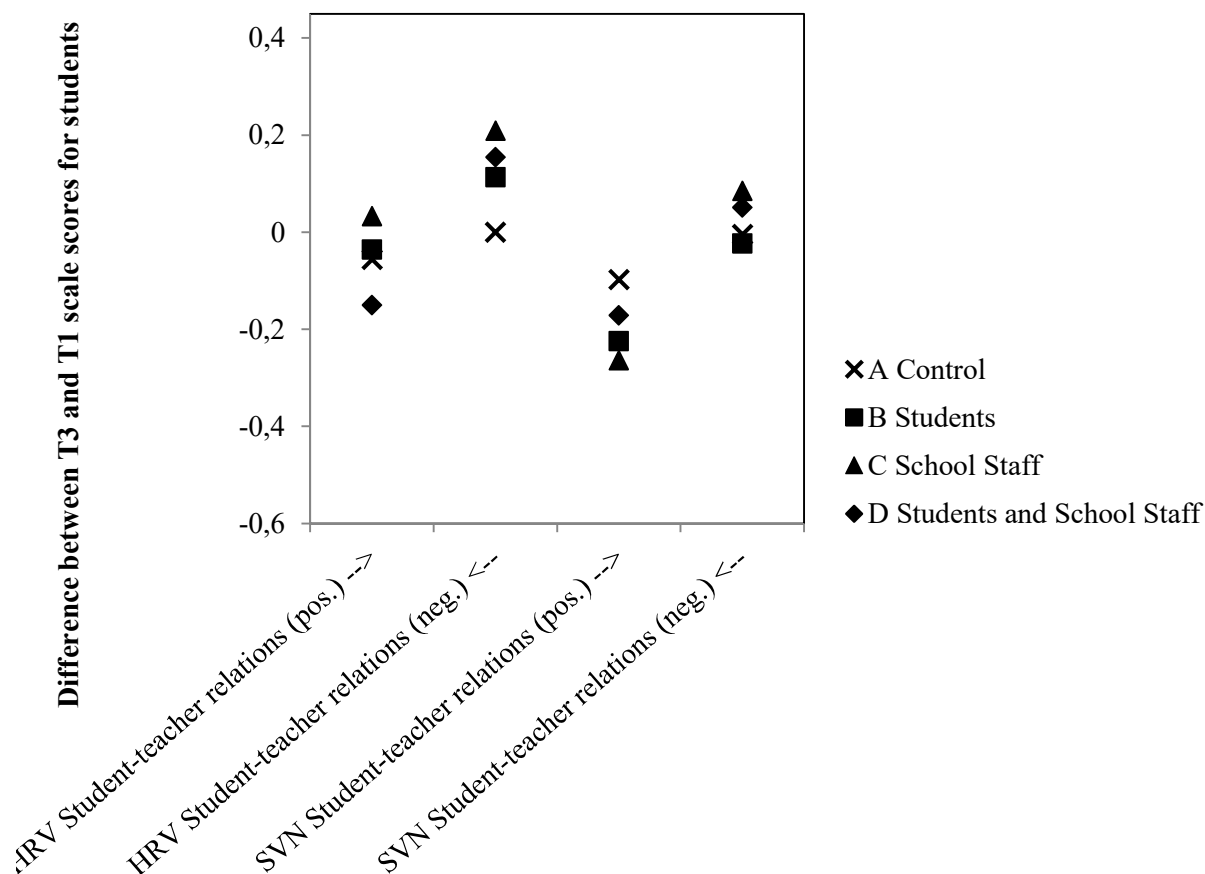


Figure 7.6. Average difference between T3 and T1 per group for student reports of the quality of student-teacher relations (recognition and misrecognition of students by their teachers) in Croatia, Slovenia, and Sweden

Figures 6.5. and 6.6. show differences between experimental groups in manifest changes between T1, T2 and T3 in the quality of student-teacher relations separately for the three school systems.

- In Croatia two significant effects of the experimental groups were observed: The average change in perceived misrecognition from teachers (negative student-teacher relations) from T1 to T2 differs significantly between group D (the group where all trainings had taken place) and the control group, yet the effect sizes is very small ($t=-2.06$, $p=0.040$, $d=-0.17$). There was no change in groups B and D between the two points in time, while students in the control group as well as students in group C reported more misrecognition from teachers at T2 as compared to T1. Only the difference between group D and the control group was significant. Further, an exact opposite effect is observed when T1 and T3 are compared: Figure 7.5. shows that students' perception of teacher misrecognition in the control group does not change between T1 and T3, while there is a slight increase

in perceived misrecognition from teachers in all three experimental groups, but only the difference between the control group and experimental group D is statistically significant ($t=2.27$, $p=0.02$, $d=0.19$). However, again, the effect sizes is very small.

- In Slovenia only one significant unexpected effect of the experimental group on changes between T1 and T2 is observed: Teacher misrecognition remains the same or becomes more frequent between T1 and T2 in groups C and D, it remains the same in group B, and it becomes less frequent in the control group. The differences between groups C and D vs. the control group are statistically significant, but effect sizes are small and very small respectively (for group C: $t=2.64$, $p=0.01$, $d=0.13$; for group D: $t=7.90$, $p=0.00$, $d=0.27$). Moreover, neither this unexpected effect nor any other significant effects of the experimental group on recognition or misrecognition from teachers are found when the manifest differences between T1 and T3 are compared.
- In Sweden no effect of the experimental group on the manifest difference between T1 and T2 regarding students' reports of recognition and misrecognition from teachers is observed.

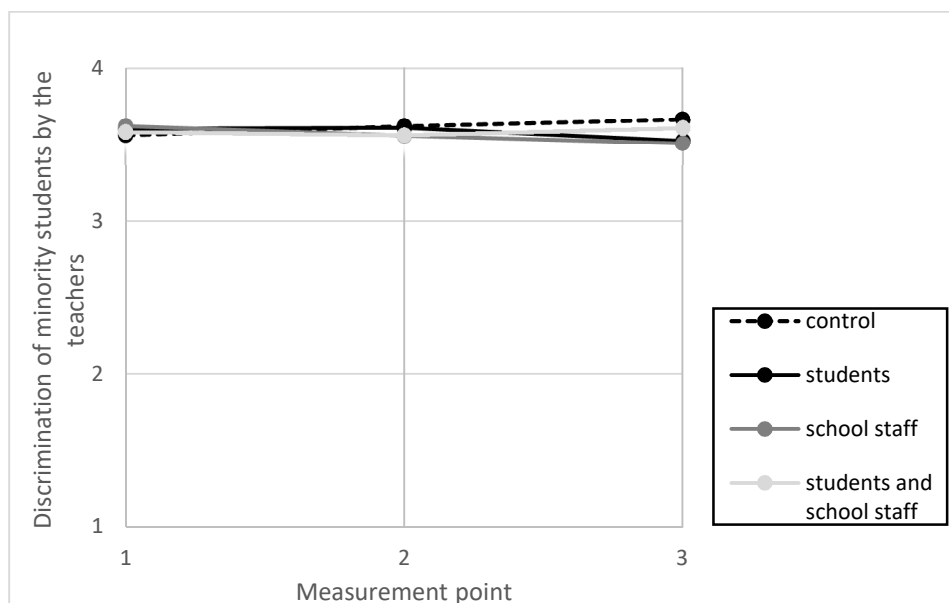


Figure 7.7. Changes in teachers' discrimination of minority students during the school year in the four experimental groups

With regard to students' perception of teachers' discrimination of minority students, Figure 7.7. shows that, on average, students think that only few of their teachers discriminate against minority students and this did not change during the school year (effect of measurement point: $F=0.13$, $df=1$, $p=0.88$). There is further neither a significant difference between experimental groups in overall means

($F=1.12$, $df=3$, $p=0.36$) nor a significant interaction between measurement point and experimental group ($F=0.10$, $df=3$, $p=0.96$). Hence, participation in the HAND in HAND programme had no effect on students' assessment of the degree to which teachers discriminated against minority students.

In summary, there is no evidence for a long-term effect of any of the HAND in HAND programmes on the quality of student-teacher relations.

7.4. Changes in the disciplinary classroom climates

Students and teachers were both asked to assess the disciplinary climate in the target class. Students were asked about their perception of the orderliness of the classroom, teachers about the frequency of unproductive student behaviours during lessons. Both scales concern classroom-level processes and, hence, changes in both were analysed at the classroom level with aggregated data. Results are presented by Figures 6.8. and 6.9.

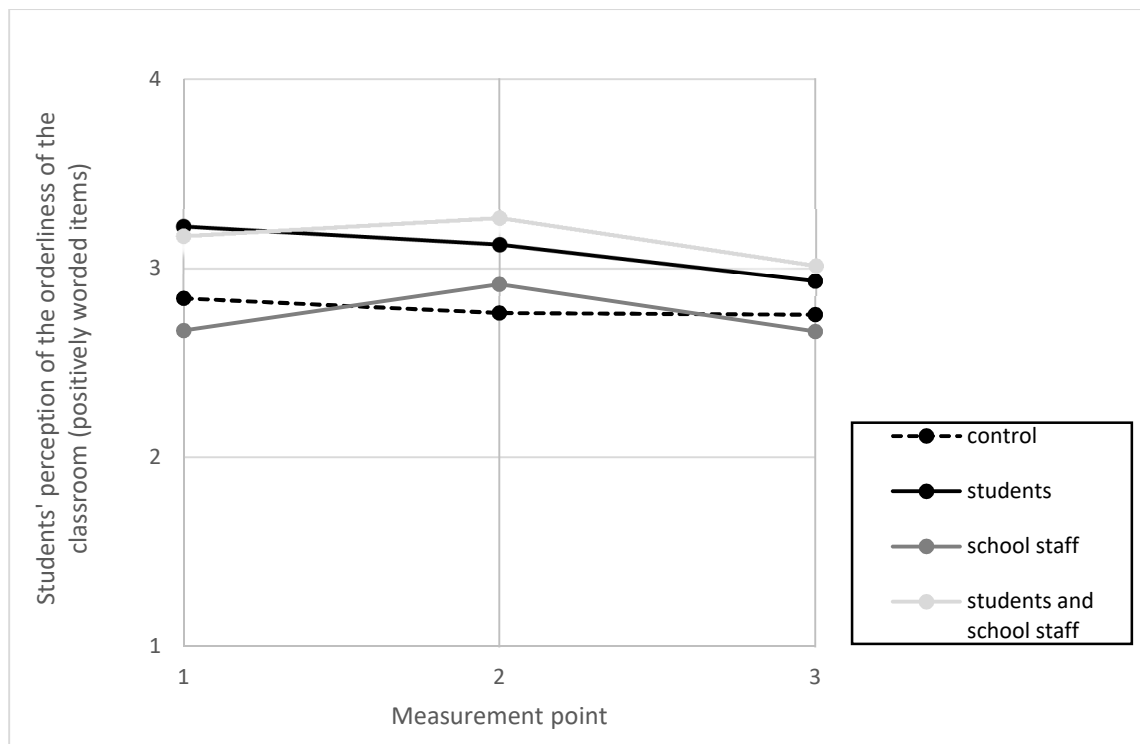


Figure 7.8. Changes in the orderliness of the classroom (negatively worded) during the school year in the four experimental groups

Figure 7.8. shows that students evaluated the orderliness of their classes as rather high and no difference between experimental groups was observed in this regard (the main effect of the experimental group was not significant: $F=1.13$, $df=3$, $p=0.36$). The quality of the disciplinary climate did further not significantly change between measurement points T1 (beginning of the school year) and T2

(after the HAND in HAND programme). It slightly worsened between measurement points T2 (after the HAND in HAND programme) and T3 (follow-up half a year later). This effect of the measurement point is significant and the effect size is large ($F=4.41$, $df=1$, $p=0.02$, $\eta^2=.14$). Yet, the worsening of the disciplinary climate is found in all experimental groups in a similar way: the interaction between measurement point and experimental group is not significant ($F=2.02$, $df=3$, $p=0.13$). This means that the observed changes in orderliness of the classroom cannot be attributed to the experimental intervention.

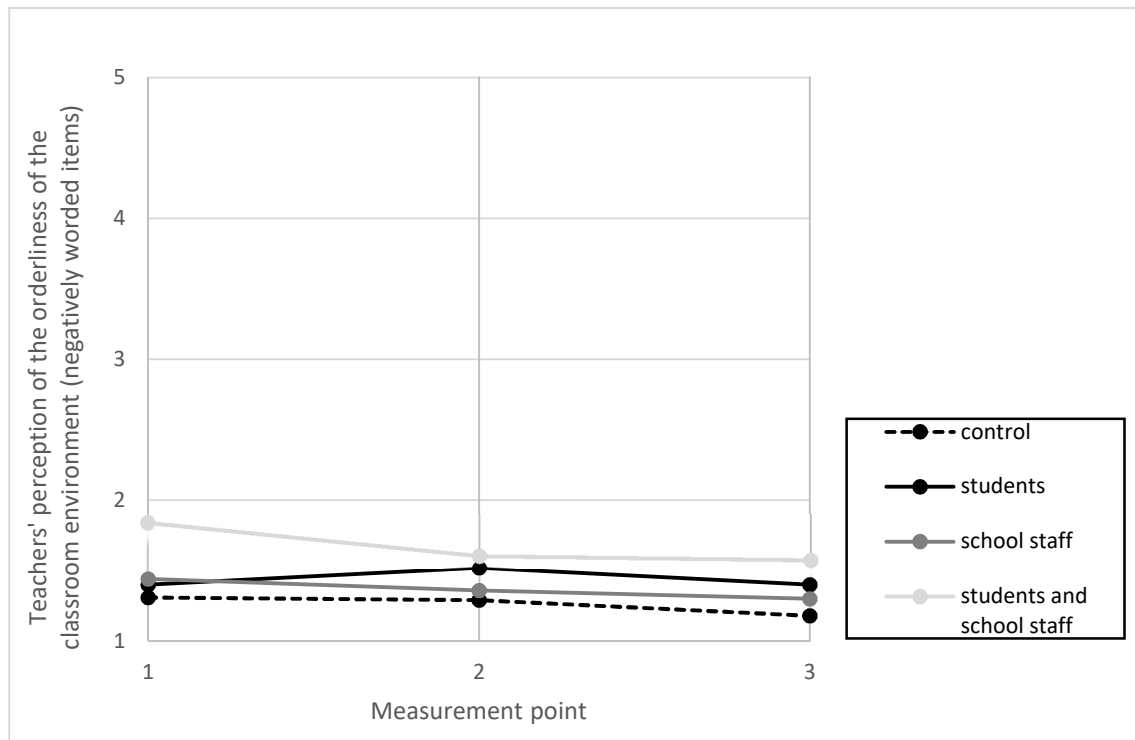


Figure 7.9. Changes in teachers' perception of the orderliness of the classroom (negatively worded) during the school year in the four experimental groups

Teachers reported that unproductive student behaviours, such as being late for class, disrupting the lesson, making impertinent remarks, etc., do occur in the target classes, but not permanently (see Figure 7.9.). The effect of the measurement point on teachers' reports of the disciplinary classroom climate was not significant ($F=4.00$, $df=1$, $p=.06$). However, a significant and strong main effect of the experimental group was observed ($F=3.80$, $df=3$, $p=.02$, $\eta^2=.29$). Posthoc tests show that teachers at schools that participated in both, the HAND in HAND student programme and the HAND in HAND school staff programmes, reported a higher frequency of unproductive student behaviours than teachers in the control group. No such difference between experimental groups was found when students' reports of disciplinary climates were considered (see above). Arguably, teachers' reports might have a higher validity in this regard, because

students lack a comparison standard (most of them have not experienced lessons in other classes). In any case, there was no significant interaction effect between the experimental group and the measurement point ($F=.63$, $df=3$, $p=.61$). Hence, the results suggest that classes that participated in student and school staff programmes might have had a poorer disciplinary climate at measurement point T1, but this did not change through participation in the programme.

In summary, there is no evidence of a positive effect of participation in the HAND in HAND programmes on the disciplinary classroom climate – no matter whether students’ or teachers’ perceptions of the disciplinary classroom climate are considered.

7.5. Results based on the sociometric measure

Sociometry was used to analyse, whether the pattern of social relations in the classroom changed as a result of the HAND in HAND programmes. Students were asked with whom they had most commonly spent their break/recess during the past 4 months. The three indicators computed based on students’ responses to this question – “density of the classroom network”, “number of students who feel isolated” and “number of unpopular students” – are all classroom-level descriptors, hence, variance analysis for repeated measures was carried out at the classroom level. However, this was only possible for 23 schools in 2 school systems, because in Sweden the sociometry was not approved by the ministry. Results are presented by figures 6.10., 6.11. and 6.12.

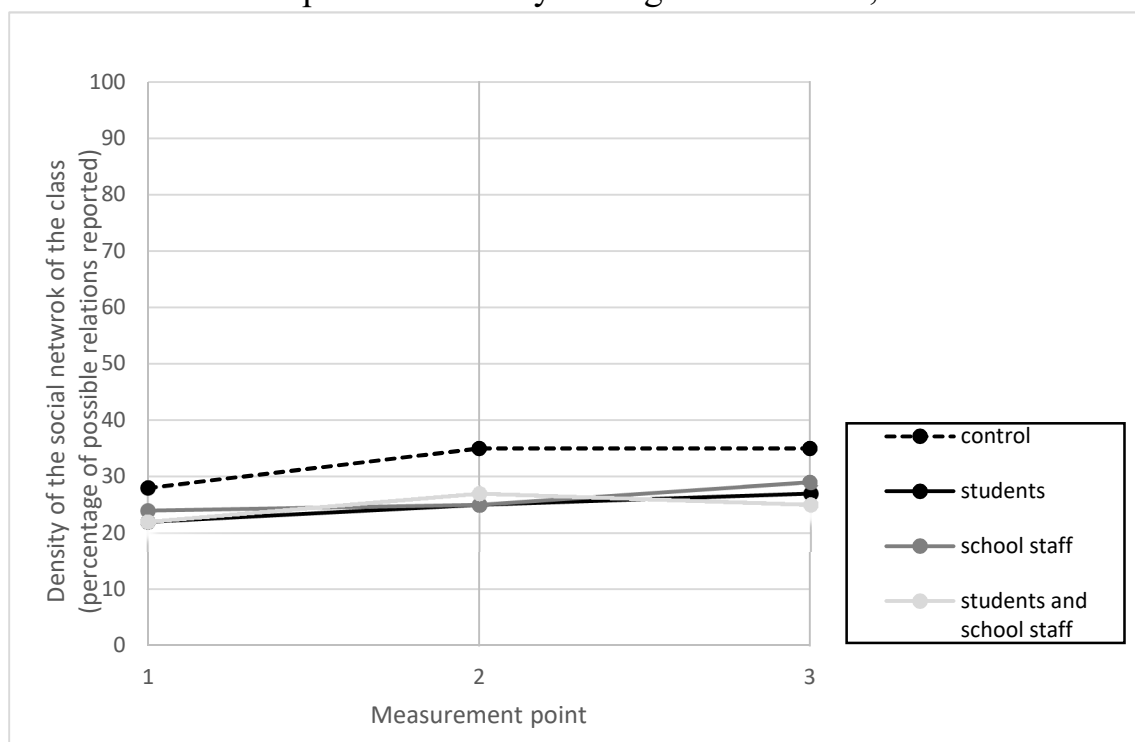


Figure 7.10. Changes in the density of the social network of the class during the school year in the four experimental groups

Figure 7.10. suggests that the density was higher in classes belonging to the control group as compared to the three experimental groups, however, this difference was not significant ($F=3.71$, $df=1$, $p=.07$). The effect of the measurement point is significant and strong ($F=7.29$, $df=1$, $p=.01$, *partielles* $\eta^2=0.28$), but the interaction between measurement point and experimental group is not significant ($F=0.21$, $df=3$, $p=.89$) and this is also the case when only two groups are compared; the control group vs. the three other experimental groups pooled ($F=0.53$, $df=1$, $p=.48$). Hence, the HAND in HAND trainings do not appear to have an effect on the density of the student-network within a class.

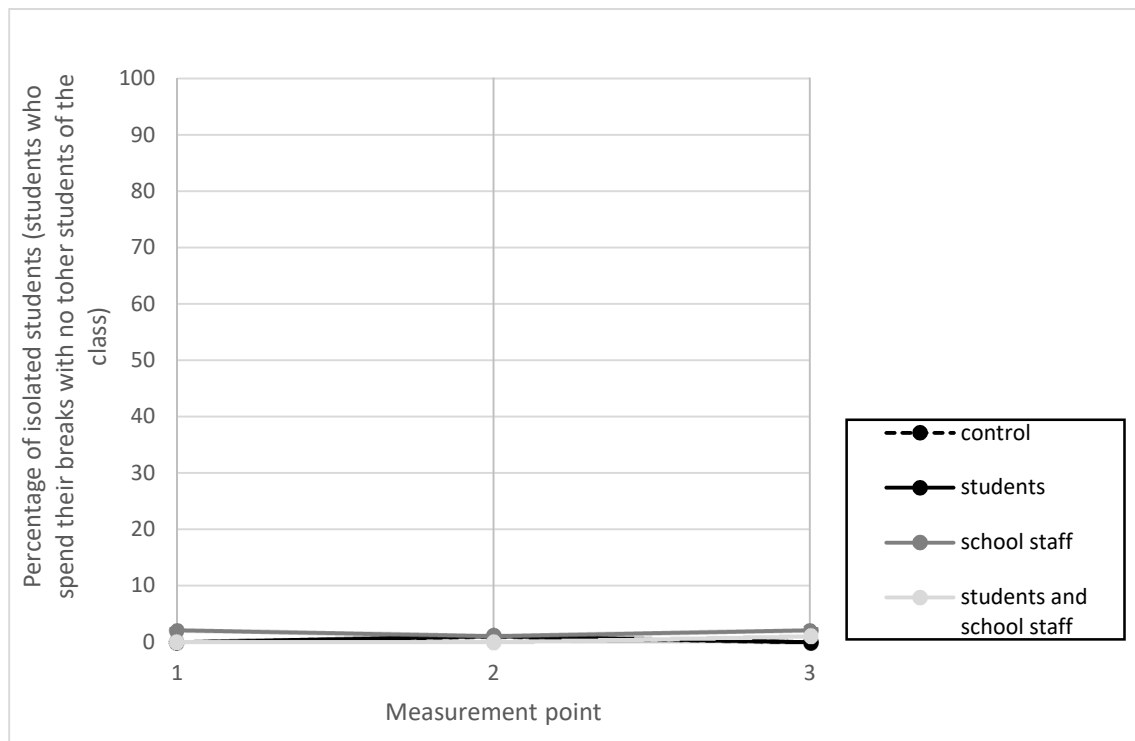


Figure 7.10. Changes in the number of isolated students (students who report that they spent their breaks with no other student from the class during the past four months) during the school year in the four experimental groups

Across classes there were only few students who felt isolated at T1, as figure 11 shows. Neither the difference between experimental groups ($F=1.44$, $df=3$, $p=.26$) nor the effect of the measurement point was significant ($F=0.09$, $df=1$, $p=.77$). More importantly, there was no significant interaction between both ($F=0.21$, $df=3$, $p=.89$), and this is also the case when only two groups are compared; the control group vs. the three other experimental groups pooled ($F=0.04$, $df=1$, $p=.85$). Hence, there is no evidence of an effect of the HAND in HAND

programmes on the number of students, who does not spend breaks with other students in a class.

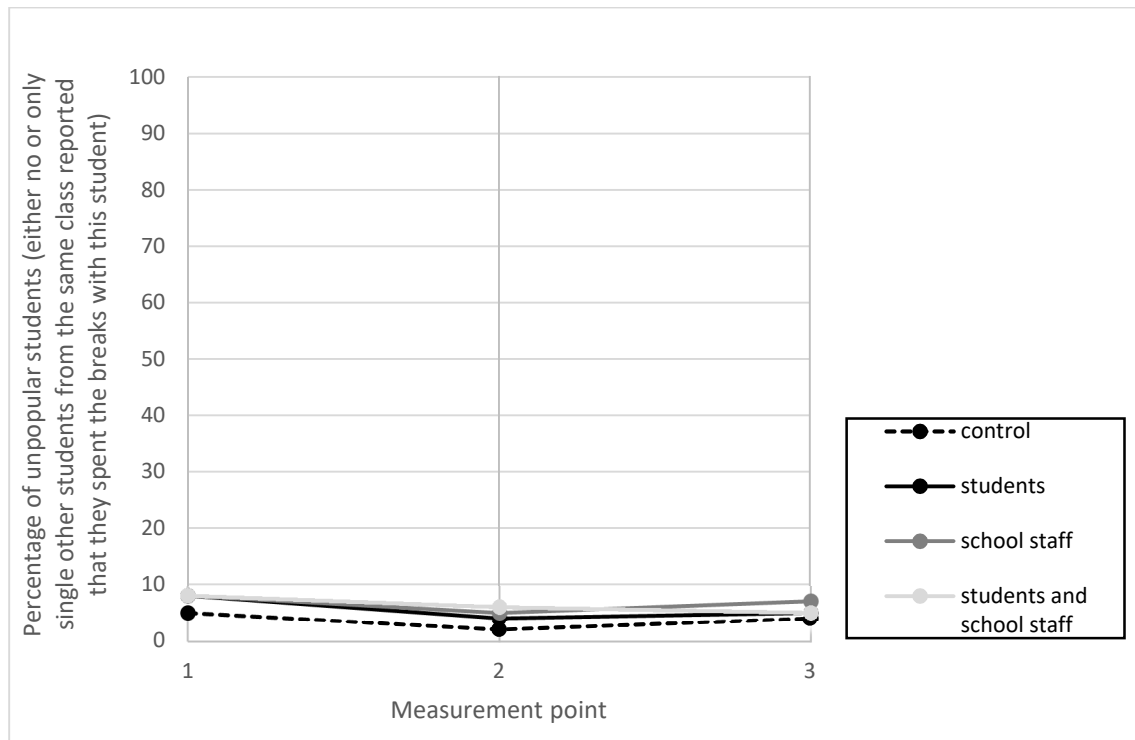


Figure 7.11. Changes in the number of unpopular students (either no or only single other students from their class reported that they spent their breaks with this student during the past four months) during the school year in the four experimental groups

In the classes participating in the HAND in HAND study there were also only few students who were unpopular in the sense that none or only one of the other students reported to spend breaks with them during the past 4 months (see Figure 7.11.). Neither the difference between experimental groups ($F=1.21$, $df=1$, $p=.33$) nor the effect of the measurement point groups ($F=2.43$, $df=1$, $p=.14$) was significant. More importantly, there was no significant interaction between both groups ($F=0.19$, $df=3$, $p=.90$). Hence, there is no evidence of an effect of the HAND in HAND programmes on the number of students, with whom only one or none of the other students has spent his*her breaks.

In summary, participation in the HAND in HAND programmes does neither appear to affect the density of the student network in class, nor the number of students who have no or few contacts with classmates during breaks.

7.6. Results from semi-structured focus-group interviews

During the semi-structured focus group interviews students and teachers were asked several questions about their perception of changes in classroom climates.

Teachers as well as school leaders and other school staff were further asked about their perception of changes in the quality of relations among school staff (which is one aspect of the school climate).

7.6.1. Which changes in classroom climates did students and teachers, observe?

To assess changes in classroom climates, students were asked how they think students behaved in their classroom and whether they had noticed any changes in the way fellow students behaved since the beginning of the school year. They were further asked how they would describe the relationships among students in the classroom and whether they noticed any differences in the relationships among students since the beginning of the school year. School staff were asked, how they would describe the social climate in the target class and whether they noticed any changes in the students, in the way students interacted with them or in the way students interacted with each other in the classroom since the beginning of the HAND in HAND programme. Responses were analysed with qualitative content analysis (see also Chapter 3). Results are summarized in Table 7.1.

Table 7.1.

Changes in classroom climates described by students and school staff during semi-structured focus-group interviews

	Control group	Only student training	Only staff trainings	Student and staff trainings	total
students and teachers observed a positive change in the classroom climate	2 (22%)	6 (67%)	4 (44%)	5 (56%)	17 (47%)
only students observed a positive change in the classroom climate	2 (22%)	0	1 (11%)	3 (33%)	6 (17%)
only teachers observed a positive change in the classroom climate	0	2 (22%)	1 (11%)	0	3 (8%)
neither students nor teachers observed any changes in the classroom climate	3 (33%)	1 (11%)	3 (33%)	1 (11%)	8 (22%)
students and teachers observed a negative change in the classroom climate	1 (11%)	0	0	0	1 (3%)
Neither students nor teachers responded to the interview question about changes in the classroom climate	1 (11%)	0	0	0	1 (3%)
total	9 (100%)	9 (100%)	9 (100%)	9 (100%)	36 (100%)

Table 7.1. shows that students and teachers agreed in several schools (17, i.e. 47%) that the classroom climate of the target class had changed in a positive way between the beginning of the school year and the time of the interview (which took place a while after the HAND in HAND training had been completed). A positive change was particularly often reported for schools/classes where either a HAND in HAND student and/or a HAND in HAND staff training had been implemented (67% of the schools/classes where only a student training had been implemented, 44% of the schools/classes where only staff trainings had been implemented and 56% of the schools/classes where student and staff trainings had been implemented). However, also in 25% of the schools/classes that had been allocated to the control group did students and teacher unanimously perceive a

positive change to the classroom climate (see Table 7.1.). Unanimous negative assessment of changes to the classroom climate concerned only one school/class and this belonged to the control group. In 11% of the schools where only a student training had been implemented, in 33% of the schools where only staff trainings had been implemented and in 11% of the schools where all three trainings had been implemented did neither students nor teachers perceive any changes to the classroom climate. In the control group, this also pertained to 33% of the classes. Finally, disagreement between students and teachers was observed in a few schools (altogether 25%). Hence, it seems that slightly more positive changes and less negative changes to classroom climates have taken place in schools where one or several of the HAND in HAND trainings had been implemented as compared to the control group. However, this difference is not significant ($Chi^2=14.17$, $df=12$, $p=.29$).

What types of changes to the classroom climate did students and teachers in different experimental groups perceive? Many of the student groups in schools where a HAND in HAND training had taken place (6 out of 27, i.e. 22%) described during the interviews that their class had bonded since the beginning of the HAND in HAND training, that they had become closer and more connected. Even girls and boys had started to get along better, said students in 3 groups (out of the 27, i.e. 11%), and in 2 groups (out of 27, i.e. 7%) students said that there was less segregation of the class into cliques. Three groups (out of 27, i.e. 27%) further described an improvement of communication among students. They said that students did „*not interrupt each other as much anymore*“, that they „*think more about what [they] say to each other*” and that they show more mutual understanding. Two groups had also noticed an improvement of cooperation among students. Moreover, students in several of the schools/classes where a HAND in HAND training had taken place, described that the atmosphere in the class had become less conflictual after the training: In four classes (out of 27, i.e. 15%) students said that students had fewer arguments and conflicts at the time of the interview than at the beginning of the school year. In two other classes (out of 27, i.e. 7%) students felt that the class had become better at solving conflicts. In one class (out of 27, i.e. 4%) students said: „*The relationships between students who were in conflict got better after some of the workshops or activities, because they had fun together*“. Finally, one of the student groups (out of 27, i.e. 4%) said that the atmosphere in the class had become calmer, another (out of 27, i.e. 4%) that the class had become more relaxed.

Teachers' observations in classrooms where a HAND in HAND training had taken place are quite similar to those of students. They also mentioned bonding:

One group (out of 27, i.e. 4%) said that they felt that students in the class had developed closer bonds. Another group (out of 27, i.e. 4%) had observed a lessening of grouping into cliques. Two teacher groups (out of 27, i.e. 7%) further said that outsiders had become better integrated. Similar to students, an improvement of communication was also mentioned by some teachers: Two groups (out of 27, i.e. 7%) said that students had become nicer to each other and more respectful. Further one group said that students cooperated better after the training (out of 25, i.e. 4%). Also a decrease in teasing and conflicts and/or an improvement in students' conflict resolution skills were perceived by four teacher groups (out of 27, i.e. 15%). One issue that was mentioned by teachers, but not by student groups was an improvement of the disciplinary climate in the classroom: 3 teacher groups (out of 27, i.e. 11%) said that the class had become calmer and more focused on the lessons.

In classes that had been allocated to the control group different types of changes were observed by students and teachers. For example, for one of the nine control group classes, teacher interventions during the previous school year were perceived as successful by all stakeholders. Here students reported: *„In 7th grade, we had classes where we talked about relationships and we are still talking about that nowadays in our class“*. Accordingly, teachers teaching this class said: *„In the last school year, we worked regularly with this class at class hours and organized activities to improve interpersonal relationships. It was noticed that they tried to control themselves, they often succeeded, but some still didn't. They are now easier to work with, it is easier for them to follow instructions, they have fewer conflicts“*. In another class from the control group where changes in classroom climates had been observed, students reported that: *„Some of those who made some trouble have moved away from our class“*. Additionally, the teachers teaching this class said: *„Teachers were on sick leave, unstructured. Year 7 with many children with special needs. But now it has calmed down“*.

In summary, students and teachers in several classes that had participated in the HAND in HAND programmes reported an increased connectedness of students, a better integration of outsiders, an improvement in the quality of communication, less conflicts, better conflict resolution, and a better disciplinary climate in the classroom. However, these changes were noted in only some of the classes, not in all classes where a HAND in HAND training had been implemented, and not even in all of the classes where a HAND in HAND student training had been implemented. Further, some changes were also reported for control-group classes. Notable is also that some teacher groups (3 out of 27, i.e. 11%) thought that changes in the classroom climate had not been sustainable. For example, one

group said: „*The students were cooperating and talking to each other, but it did not last for long. This change was during the implementation of the HiH programme. After the programme finished, they started fighting again*“. Hence, there is no statistical evidence that the HAND in HAND training had a causal effect on classroom climates. Interview results suggest that the HAND in HAND programme might have triggered some positive changes in some classes, but they cannot be clearly attributed to participation in the training.

7.7. Which changes in the relations among school staff did teachers, school leaders and other school staff observe?

To assess changes in school climates more generally, teachers as well as school leaders and other school staff were asked, how the social climate and the cooperation among the school staff was before the training and whether they noticed any changes since the beginning of the HAND in HAND programme. Responses to these questions are summarized in Table 7.2.

Table 7.2. Changes in classroom climates described by students and school staff during semi-structured focus-group interviews

	Control group	Only student training	Only staff trainings	Student and staff trainings	total
teachers and school leaders/other school staff observed a positive change in relations among staff	0	0	3 (33%)	3 (33%)	6 (17%)
only teachers observed a positive change in relations among staff	0	1 (11%)	4 (44%)	2 (22%)	7 (19%)
Only school leaders/ school staff observed a positive change in relations among staff	0	1 (11%)	1 (11%)	1 (11%)	3 (8%)
neither teachers nor school leaders/other school staff observed changes in relations among staff	6 (66%)	5 (55%)	1 (11%)	3 (33%)	15 (42%)
teachers and school leaders/other school staff observed a negative change in relations among staff	0	0	0	0	0
Neither teachers nor school leaders/other school staff	3 (33%)	2 (22%)	0	0	5 (14%)

responded to the question of relations among school staff					
total	9 (100%)	9 (100%)	9 (100%)	9 (100%)	36 (100%)

Table 7.2. shows that negative changes in relations among staff were reported in none of the schools and positive changes were mainly reported by school staff who had participated in a HAND in HAND staff training. More specifically, a positive change was observed by either teachers or school leaders/other school staff or both in 88% of the schools where only HAND in HAND school staff trainings had been implemented and in 66% of the schools/classes where both, a HAND in HAND student and HAND in HAND school staff trainings had been implemented. In contrast, only in 22% of the schools where only a HAND in HAND student training had been implemented did the school staff report a positive change in relations among staff and in none of the schools that belonged to the control group. The difference between the 18 schools where school staff trainings had been implemented and those 18 schools where no school staff trainings had been implemented is statistically significant ($Chi^2=12.70$, $df=3$, $p=.01$, $Cramer's V=.34$). Hence, there is indication that the HAND in HAND staff trainings had a positive effect of medium size on relations among staff.

How did the relations among staff change? All 11 teacher groups who had participated in the training and noted changes in relations among staff said that those teachers who had participated conjointly in the training had bonded, become closer and more connected due to their joint experience. In accord with the teachers, also 4 school leader/other school staff groups said that they had the impression that the staff who had participated in a HAND in HAND staff training had bonded and become closer. School leaders/ other school staff groups mentioned some additional changes: One group said that teachers cooperated better after the training and had improved their communication. Another group said: „*They laugh and have fun. Earlier they were more focused on their own subjects. Now they talk about well-being and different solutions.*” A third school leader group observed that teachers complained less than before and had started to solve more problems on their own. Finally, another school leader had the feeling that participation in the training had improved his*her relation with one specific teacher.

In both schools where a positive change in staff relations was observed by the staff even though no HAND in HAND staff training had been implemented a change in school management had taken place. The teachers in one of the schools

said: “*There have been changes since the beginning of the school year, with new school management trying to improve the relations between teachers and the atmosphere in school.*” And in the other school other staff said: “*Now with new school leadership the climate is better and implementation of new projects are encouraged now.*”

In conclusion, the results of the semi-structured focus-group interviews suggest that participation in a HAND in HAND staff training improves the relations between participants. It appears to improve their communication and cooperation.

7.7. Discussion

This chapter examined whether the HAND in HAND programmes had a positive effect on classroom and school climates. Classes and schools from the four experimental groups (control group, student programme only, staff programmes only, student and staff programmes) were compared with regard to changes in their school/classroom climates between the measurement points T1 (before the programme), T2 (after the programme), and T3 (follow-up). To increase the validity of conclusions, the chapter used a multimethod-approach: Student responses to questionnaire scales asking about the quality of student-student relations, student-teacher relations and disciplinary climates in the target class were compared to teacher responses to questionnaire scales addressing similar issues. Questionnaire results were further compared to sociometric results and to findings from semi-structured focus-group interviews, and quantitative analysis was complemented with qualitative analysis.

All in all, the findings suggest that the HAND in HAND programmes had no consistent long-term effects on the quality of classroom climates. In some of the classes students and/or teachers had observed some (mostly positive) changes to classroom climates, but not in all classes that had participated in a HAND in HAND training and also in some classes that had been allocated to the control group. Also analysis of questionnaire scales suggested that there might have been some differences between experimental groups regarding changes between T1 and T2 with regard to single scales in single school systems. Some of these were in the expected direction (positive effects of the trainings), but others were in the unexpected direction (negative effects of the trainings). However, no effects of experimental groups on changes between T1 and T3 were observed – neither when students’ assessments of classroom climates were considered nor when teachers’ assessments of classroom climates were the focus. Additionally no changes to the sociometric structure of the classroom were found – neither did the density increase nor did the number of isolated and/or unpopular students decrease

after participation in a HAND in HAND programme. Hence, there is no clear evidence that any of the HAND in HAND programmes improved classroom climates in the participating classes. Possibly, the programme might have had such an effect in specific classes only – for example in classes that were particularly interested and willing to engage. Interview results point in that direction. However, future research with larger samples would be needed to test this assumption, as the sample for the present evaluation was too small to examine moderator effects of classroom characteristics.

Unlike the non-significant effects of the experimental condition on classroom climates, the effect of the experimental condition on the quality of relations among staff in the school was statistically significant and the effect size was medium. Positive changes to relations among staff were reported by staff significantly more often in schools where the staff had participated in a HAND in HAND school staff training as compared to the other schools (those that either belonged to the control group or where only a student training had taken place). Accordingly, interviews with school staff suggest that the school staff who had participated in the training had bonded as a result of their common experience. They had become closer and cooperated better after the training. At least, this is a good fundament for future cooperation among staff in reflecting their teaching, and it can be considered a very first step for improving the overall school climate.

7.8. References

- Adelman, H. S., & Taylor, L. (2002). Classroom climate. In S. W. Lee, P. A. Lowe, & E. Robinson (eds.), *Encyclopedia of school psychology* (pp. 304-312). Thousand Oaks, CA: Sage. Pp. 304–312.
- CASEL (2013). URL: <https://casel.org/what-is-sel/>
- Cohen, J., McCabe, E., M., Michelli, N. M., & Pickeral, T. (2009). School Climate: Research, Policy, Teacher Education and Practice. *Teachers College Record, 111*(1), 180-213.
- Davis, H. A. (2003). Conceptualizing the Role and Influence of Student-Teacher Relationships on Children's Social and Cognitive Development. *Educational Psychologist, 38*(4), 207-234.
- Hennig, M., Brandes, U., Pfeffer, J., & Mergel, I. (2012). *Studying Social Networks : A Guide to Empirical Research*. Frankfurt [u.a.]: Campus Verlag.

- Jansen, D. (2006). Einführung in die Netzwerkanalyse: Grundlagen, Methoden, Forschungsbeispiele (3. Auflage). Bielefeld: VS Verlag.
- Moos, R. H. (1979). *Evaluating educational environments*. San Francisco, CA: Jossey-Bass.
- Rosenthal, R. (1994). Interpersonal Expectancy Effects: A 30-Year Perspective. *Current Directions in Psychological Science*, 3(6), 176-179.
- Van Houtte, M. (2005). Climate or Culture? A Plea for Conceptual Clarity in School Effectiveness Research. *School Effectiveness and School Improvement*, 16(1), 71-89.
- Walberg, H., & Anderson, G. (1968). Classroom climate and individual learning. *Journal of Educational Psychology*, 59, 414 – 419.
- Wilson, H. K., Pianta, R. C., & Stuhlman, M. (2007). Typical classroom experiences in First Grade: the role of classroom climate and functional risk in the development of social competencies. *The Elementary School Journal*, 108, 81–96.

**Chapter 8:
Participants' summative evaluation of the HAND in HAND
programmes¹⁹**

Svenja Vieluf, Nina Roczen, Mojca Rožman

¹⁹ Please note that part of the analysis presented in this chapter has already been published in Vieluf, Denk, Rožman and Roczen (2020).

8.1. Introduction

The present chapter addresses the question, how participants themselves evaluated the HAND in HAND programme. It complements findings from the experimental control-group study, presented in Chapters 5 and 6. It helps to understand the perspective of those persons the programme is intended to serve. First, the chapter informs about what participants think has changed as a consequence of the HAND in HAND programme: Did they observe those changes that have been revealed through the repeated use of questionnaires, tests, and sociometric procedures in the experimental control-group study? Did they notice additional effects and, if yes, are these intended or unintended, positive or negative? Second, the chapter also informs about another important question: Did participants like the programme and what did they like/ dislike in particular? Findings are based on responses to semi-structured focus-group interviews (see also Chapter 3).

8.2. What do participants think were the main effects of the HAND in HAND programme?

To find out what participants think were the main effects of the HAND in HAND programme, responses to the following three questions asked during the semi-structured focus-group interviews are considered: 1. Through the HAND in HAND training, what did you learn? (question asked during the semi-structured focus-group interviews with students) 2. What would you highlight as perceived outcomes from participating in the training activities? (question asked during the semi-structured focus-group interviews with teachers as well as in those with school leaders/other school staff) and 3. What did you personally learn from the trainings? (question asked during the semi-structured focus-group interviews with teachers as well as in those with school leaders/other school staff)²⁰. For analysing the responses qualitative content analysis was used (Schreier, 2012, see also Chapter 3). However, the definition of categories for the first part of the analysis was theory-driven, not data-driven. We coded whether changes in dimensions of the CASEL-model, in intercultural competencies/diversity awareness, or in school climates (see Chapter 1 for a more detailed description of intended outcomes) were mentioned by participants. Additionally, we coded whether other outcomes were noted by the participants, i.e., outcomes that were not explicitly intended by the HAND in HAND programme. We formed inductive lower-order

²⁰ Later during the semi-structured interviews participants were also asked, whether they think that the classroom and school climates have changed as a consequence of the HAND in HAND programme. Responses to this question were discussed in Chapter 12 and are not considered here, because, here, we are exclusively interested in effects that participants come up with spontaneously and without suggestion of the interviewer

categories for these “other” responses. The results are further described separately for the three stakeholders: students, teachers, and school leaders/other school staff²¹.

8.2.1. Students

Three student groups (out of 18, i.e. 17%) said that they had learned “nothing” (2 groups) or “not much” (1 group) from the HAND in HAND programme. All other student groups mentioned some of the intended outcomes in their responses to the question, what they learned through the HAND in HAND programme. Most frequently they mentioned intercultural competencies/diversity awareness, closely followed by self-management competencies and relationship skills. Self-awareness, social awareness and responsible decision making were also mentioned, but only by a few groups. Results are described in more detail in the following:

- ***Intercultural competencies/diversity awareness*** was mentioned by seven student groups (out of 18, i.e. 40%) during the semi-structured focus group interviews. In particular, several students said that they learned diversity-related norms during the HAND in HAND programme, i.e., that “*one has to respect everyone regardless of religion, nation and the like*”, that “*social discrimination is a bad thing*”, that “*we should not have prejudices, we should not be racist*”, or that “*we should not judge others by their looks*”. Some students further said they gained knowledge about stereotypes, prejudices or tolerance. Other students said that HAND in HAND created awareness for social inequalities and for the difficulties some minorities face in society. In one group were students who said that they actually learned to behaviourally address diversity. They said they now know how to “*behave towards people that are different from the majority and how to be tolerant towards them*” and “*how to approach people we don't know without prejudice*”.
- ***Self-management competencies*** were mentioned by six groups (out of 18, i.e. 33%). One group said: “*If I'm angry, then I think I'm happy and then it goes away*”. Similarly, another group said: “*If we think positively that we are in a better mood*”. Students in the other groups said they learned to think before they act, to control their own emotions, to relax and calm down, and to “*deal with difficult emotions such as loneliness*”. Two students (from different interview groups) explicitly stated that they had

21 »Other school staff« encompasses i.e. the student health teams, special needs teachers, school social workers, school counsellors, school psychologists and similar professions.

successfully used inner exercises to calm down in situations where they had been nervous.

- **Relationship competencies** were mentioned by 6 groups (out of 18, i.e. 33%). Students said they learned “*social skills*” or “*how to better communicate with others*”, not to interrupt each other and to “*listen to both sides of the story*”, to “*share feelings with others*”, to cooperate and/or to deal with conflicts and violence.
- **Self-awareness** was mentioned by two groups (out of 18, i.e., 11%). Students in one of these groups said that they learned how to “*interpret the own emotions*”. In the second group a student said that he*she got more confidence.
- **Social awareness** was mentioned by two groups (out of 18, i.e. 11%). Students in these groups said they learned to “*step in other person’s shoes*” or to “*take the perspective of others*”.
- **Responsible decision-making** was mentioned by only one group (out of 18, i.e. 6%). In this group one student said they learned to deal with problems.

In addition to Social and emotional competencies and intercultural competencies/diversity awareness, also improvements of the **classroom climate** were mentioned by four groups (out of 18, i.e. 22%) in response to the question, what they have learned from the programme, two groups said that the students in the class had become closer. Another group said that the class became “*more united as a group*”. The fourth group said that they “*connected more, talk more and like each other more*”.

Some responses of students were not specific enough to map them on the intended outcome-dimensions, but still addressed related issues: Two of the 18 groups (i.e. 11%) said that the HAND in HAND programme for students had the effect to “*raise awareness*” or that it was an “*eye-opener*”. One of the groups (i.e. 6%) further said that they now look at themselves and others differently. Two groups (out of 18, i.e. 11%) said that they had the impression that the programme helped shy students in particular. No unintended effects were mentioned by students.

8.2.2. Teachers

Asked what they would highlight as perceived outcomes from participating in the HAND in HAND training, teachers referred to several competencies that the programme had aimed to foster. In particular, self-management was mentioned frequently, closely followed by relationship skills and self-awareness. Single groups also mentioned social awareness or intercultural competencies/diversity

awareness. Responsible decision making was not mentioned. The results are described in more detail in the following:

- **Self-management** was mentioned by seven groups (out of 18, i.e. 39%). In three of these groups teachers said that they became aware of the importance of taking care of their own well-being. In four of the seven groups teachers said they learned to cope with stress. Also four groups said they learned to stay calm in difficult situations (including situations where students try to provoke them) and not to react impulsively. Two groups said that they learned to take things not too personal and to emotionally distance themselves from the situation when necessary. For example, one of these latter groups said that *“it is important to realize that conflict situations are not directed at us personally, but at the role of the teacher that we represent”*.
- **Relationship skills** were mentioned by six groups (out of 18, i.e., 33%). One of these groups learned to avoid getting personally involved in a conflict or problematic situation in school. Another group said that they learned to deal in a better way with difficult situations in the classroom. A third group said they learned *“how to approach professionally conflict situations that occur regularly in schools”*. In a third group one teacher said: *“Personally, I notice at myself that I started to open and connect with others.”* In the fifth group two teachers said that they had learned to say *“no”*. Finally, in one group a teacher said, he*she learned *“that emotions infect. If we stand for the norm, it affects the students. If we are calm, then we get calm students”*.
- **Self-awareness** was mentioned by two groups (out of 18, i.e. 11%). In one of these groups a teacher said that he*she perceived as an outcome of the HAND in HAND programme: *“becoming more aware of oneself, reflecting on oneself”*. In another group a teacher said that he*she now thinks about how he*she breathes.
- **Social awareness** was mentioned by one group (out of 18, i.e., 6%). One teacher in this group said that he*she now *“listen[s] to students more and consider[s] their needs for breaks.”*
- **Intercultural competencies/diversity awareness** was mentioned by one group (out of 18, i.e. 6%). Teachers in this group said that they have become more *“aware of social and cultural differences among people and in society”*.
- **Responsible decision making** was not mentioned by any of the teacher groups in response to the relevant interview questions.

An improvement of *the school climate* was mentioned by two groups (out of 18, i.e. 11%). Teachers in these groups said that they have bonded with colleagues and one of the groups also said that they will cooperate better in the future.

Some teachers also mentioned outcomes that cannot be directly mapped on the dimensions of the CASEL-model, but were related to the aims of the programme. On a very general level, three groups said that they felt empowered by the training. Six groups further mentioned that they learned new exercises which they can use with students in the classroom or during breaks. Finally, one group (out of 18, i.e., 17%) mentioned theoretical knowledge about learning. They reported that they learned “*how oxygenation is related to learning*”. None of the teacher groups mentioned unintended outcomes.

8.2.3. School leaders/ other school staff

School leaders and other school staff also mentioned many of the aims of the HAND in HAND programme in their responses to the questions, what they would highlight as perceived outcomes from participating in the training activities and what they personally learned from the training. Most frequently they mentioned self-awareness. Self-management was also relatively frequent. Relationship skills and social awareness were mentioned by two respective one interview groups. None of the groups mentioned responsible decision-making or intercultural competencies/diversity awareness. Responses are described in more detail in the following:

- **Self-awareness:** Seven school leader/other staff groups (out of 16, i.e., 44%) mentioned self-awareness. Six of these groups said that the programme reminded them of the importance to be self-aware. For example, one group said: “*The insight that first we need to be in control and conscious of ourselves, and only after that we can act towards others.*” Additionally, one group said that they have become more focused on the self and that they do more self-reflection since the training. Another group said that they have become more aware of their emotions. Again another group said that they now feel more confident that they have the ability/competencies to confront all kinds of difficult situations in our work in schools and everyday lives.
- **Self-management** was mentioned by six interview-groups (out of 16, i.e. 25%). One group said they learned impulse-control and the other three groups mentioned stress management. Two groups (out of 16, i.e. 13%) said that they have learned about the link between body and psyche or the importance of physical well-being for psychological well-being.

- **Relationship skills** were mentioned by two groups (out of 16, i.e. 11%). Both said that they learned new communication skills.
- **Social awareness** was mentioned by only one interview-group (out of 16, i.e., 6%). A special education teacher said that he*she now experiences “*a greater understanding of children in need of support*”.
- **Responsible decision-making** was not mentioned by any of the groups.
- **Intercultural competencies/diversity awareness** was also not mentioned by any of the groups.

Three groups (out of 16, i.e. 19%) described positive changes to the **school climate** as an outcome of the training. Two of these groups said that the teachers who had participated in the training have become closer. The third groups said that they have the impression that the classroom climate in the participating class has improved; that the class has “*calmed down a little*”.

In addition, school leaders mentioned some outcomes which cannot be mapped to the CASEL-dimensions or to the construct of school/classroom climate, but concern related aspects: One group (out of 16, i.e., 6%) said they felt empowered through the programme. Two groups (out of 16, i.e. 13%) said that they have gotten to know useful activities and exercises that they can use e.g. to create active breaks. Other two groups (out of 16, i.e. 13%) said that they perceived changes in the teachers who had participated in the teacher programme. One of these groups said that the teachers appear to have more job-satisfaction now, and the other group said that they noticed a change of attitudes in one particular teacher at their school.

8.3. How did participants like the HAND in HAND programme?

This section addresses the question, how participants liked the programme and why. It is based on the following two questions asked during the semi-structured focus-group interviews: 1. How did you like the HAND in HAND programme overall? (question asked during the semi-structured focus-group interviews with students) 2. How would you evaluate the HAND in HAND programme overall? (question asked during the semi-structured focus-group interviews with teachers as well as during those with school leaders/other school staff).

8.3.1. Students

Four of the 18 student groups (22%) had a positive overall evaluation of the HAND in HAND student programme. These four groups stated the programme was “*fun*”, “*interesting*” or “*innovative*”. Eight (of the 18 groups, i.e. 44%) gave

the programme a mixed evaluation. These groups, for example, stated that “*some exercises were fun, but others were boring*”, that the programme was “*interesting*” but that they “*had expected more*”, or that the programme was “*okay*”. A mixed evaluation might also imply that some students within the interview group found the programme better than others. In 1 of the 18 student groups (6%), students had a largely negative evaluation. This group called the programme “*childish*”, “*not serious enough*” and “*boring*”. Finally, 4 of the 18 groups (22%) only made specific comments and did not provide an overall evaluation. Hence, the students' evaluations were mostly mixed. Many students found the programme ‘ok’, but were not enthusiastic about it. Still, different students liked the programme better than others and there were also a few students who expressed dislike.

When students gave reasons for their positive evaluations, they often argued that the programme had been “*fun*” (5 out of 18, i.e. 28%) or referred to the programmes' “*interestingness*” (4 out of 18, i.e. 22%). Further, 3 groups (out of 18, i.e. 17%) argued that the programme was, at least, better than regular lessons. Related to this, one group liked the fact there was no need to sit still during the programme, that they had the opportunity to express their opinion, that they did not have to study, and that they were not given grades. Two groups (out of 18, i.e. 11%) appreciated that the programme had helped with connecting with classmates and two groups (out of 18, i.e. 11%) said that it was relaxing. Finally, one Croatian group (out of 18, i.e. 11%) stated: “*There are many things that can be learned from the HAND in HAND programme, like how to deal with conflicts and violence, about emotions like loneliness and others*”.

Negative evaluations sometimes referred to the “*boringness*” of individual exercises (4 out of 18, i.e. 22%). Three groups (out of 18, i.e. 17%) believed the programme was not adapted to their level of maturity and knowledge, i.e. that the exercises were “*childish*” or that they already “*knew many of these things that were taught during the programme*”.

8.3.2. Teachers

The teacher evaluations of the programme were considerably more positive than those of the students. Fourteen out of 18 teacher groups (i.e. 78%) agreed on a positive overall evaluation of the HAND in HAND teacher programme. These groups called the programme “*good*”, “*interesting*”, “*useful*”, “*an excellent experience*” or “*the best training so far*”. One group said that it “*worked well*”, another that they “*liked*” the programme or “*really enjoyed*” it. Four out of 18 groups (22%) gave the programme a mixed evaluation. Teachers in these groups said positive things in response to the question of how they evaluated the

programme overall, but also voiced some criticism. No evaluation was clearly negative.

In response to the question about their overall evaluation of the programme, teachers mentioned a variety of evaluation criteria. First of all, the majority of teacher groups (11 out of 18, i.e. 61%) substantiated their positive evaluations of the HAND in HAND programme with their liking of its specific focus. In fact, many teacher groups found this focus quite unusual (7 out of 18 groups mentioned this, i.e. 39%), in particular, teachers in Croatia. For instance, one teacher stated that it had been the first programme that was *“focused primarily on the empowerment of the teachers”*. Another teacher said: *“It really seemed important to me that the focus was also on the teacher”*. It is noticeable that something about the programme made some teachers feel recognised in a way that was apparently exceptional, as best illustrated by the following quote: *“That was what I liked about this programme: One felt important”*.

Apart from the focus, many teacher groups also commented positively on the content and design of the HAND in HAND programme: Half the teacher groups (9 out of 18, i.e. 50%) mentioned the exercises were useful. For example, one teacher said: *“it was applicable in the classroom, but we also learnt the techniques for self-awareness and personal growth which we as teachers need as it is a stressful job”*. Another teacher stated: *“There were new exercises that felt applicable to the school and that can be used in everyday life”*. In addition, 3 groups (out of 18, i.e., 17%) emphasised that they had become familiar with a variety of exercises. An example statement is: *“It was a good mix of exercises, group strengthening, physical, relaxation. A good package”*. Nine teacher groups (out of 18, i.e. 50%) also liked that the programme had a hands-on approach. Three groups (out of 18, i.e. 17%) positively viewed the programme for not being presented as a fixed sequence of exercises, but as a fund of ideas, exercises and techniques from which they could choose whichever seemed most suitable for their situation and purpose.

The implementation of the programme also attracted many positive comments: Seven teacher groups (out of 18, i.e. 39%) mentioned they appreciated the positive atmosphere during the programme. For example, one teacher said: *“The atmosphere was relaxed, we were very relaxed at the programme too”*. Another stated: *“It was a good feeling and an atmosphere of acceptance”*. As these quotes indicate, several teacher groups also said the programme made them feel good (all 6 Slovenian groups). For example, one teacher said: *“We were having a rest, we laughed, and we were full of energy”*. Another teacher stated: *“I always came back*

in a better mood than when I had left". Three teacher groups (out of 18, i.e. 17%) supported their positive evaluation of the programme by referring to its positive effect on the social relationships between the participating teachers. Four groups (out of 18, i.e. 22%) spontaneously praised the trainers. Finally, three groups (out of 18, i.e. 17%) liked staying in a hotel.

Negative aspects mentioned in response to the question about overall evaluation of the programme were that teachers found implementing the exercises in their own classrooms difficult (5 out of 18 groups mentioned this, i.e. 28%), that it was exhausting to participate for 6 hours in a programme session after a long working day on Thursday and Friday afternoons (2 out of 18 groups mentioned this, i.e. 11%), and that the exercises started repeating after the first session (1 out of 18 groups, i.e. 6%).

8.3.3. School leaders/other school staff

School leaders and other school staff also held largely positive evaluations of the HAND in HAND programme. Eleven school leader/other school staff groups (out of 16, i.e. 69%) had a positive overall evaluation of the HAND in HAND programme. For example, these groups called the programme "*interesting*", "*useful*", "*helpful*" or "*successful*". One group even said they were "*impressed by the programme*". Five (out of 16 groups, i.e. 31%) only gave the programme a mixed evaluation. Here, school leaders and/or other school staff said some positive things in response to the question of how they evaluated the programme overall, but also expressed some criticism.

The school leaders mentioned a variety of evaluation criteria in their responses that were partly different from those mentioned by the teachers and students. Similar to the teachers, many school leader/other school staff groups generally liked the focus of the programme. Five groups commented positively on the focus on personal growth and empowerment (out of 16, i.e. 31%) and three groups commented positively on the focus on relationship-building (out of 16, i.e. 19%). Four groups (out of 16, i.e. 25%, all 4 in Croatia) described these foci as being "*new*" and "*different*" from other programmes.

The content and design of the HAND in HAND programme as well as its implementation was mentioned less often by the school leaders/other school staff than by the teachers: Two groups (out of 16, i.e. 13%) said they enjoyed the programme and had had a good time. Three groups (out of 16, i.e. 19%) praised the experiential approach of the programme. Individual groups also liked the whole-school-approach, that there was no time pressure during the programme,

the possibility of exchanging with colleagues from other schools, and that the programme did not hinder the school process because it took place outside of lesson time. Another aspect mentioned by one group was the European dimension of the HAND in HAND project.

Instead of praising the content and/or implementation of their own programme, the school leaders/other school staff often commented positively on the teacher programme. Six groups (out of 16, i.e. 38%) expressed that the teachers liked their programme and two groups (out 16, i.e. 13%) said they thought the programme had a positive effect on teachers. For example, one principal said: *“It seems that they felt that they were helped by HAND in HAND. It suited the lessons well”*. Another said: *“Teachers think a little different now; they have done some exercises in class”*.

The main criticism voiced by school leaders and other school staff was that their own programme had been too short (only 2 days): six groups mentioned this (out of 16, i.e. 38%). For example, one group said the programme had been *“an initial spark, but it remained somehow unfinished”*. Three groups (out of 16, i.e. 19%) were further critical of having been separated from the teachers in the programme. One group complained they did not even get to know what the teachers and students had done in their programme. In one group in Sweden, the student health team and teachers for special needs also said they had felt left aside. They suggested: *“We could be the motors instead of testing without practising before”*. And one group said they found it a pity that it was not possible to let the whole staff at the school participate.

Finally, stronger criticism of the HAND in HAND programme comes from three school leader/other school staff groups: Two of these groups stated they doubted the applicability of the HAND in HAND exercises in the classroom (1 Slovenian, 1 Swedish) and one Swedish group doubted whether the programme had any effects.

8.4. Summary and Discussion

This chapter aimed at answering two questions: Which effects of the HAND in HAND programme did participants observe? And: How did they like the HAND in HAND programme overall and why? Results from the interviews concerning these two questions concern the quality of the programme and, thus, contribute to its summative outcome evaluation. They are summarized and discussed in the following.

8.4.1. *What do participants think were the main effects of the HAND in HAND programme?*

Participants described a variety of outcomes of the HAND in HAND programme. Most of these were congruent with the programme's aims. Many participants said that they had acquired Social and emotional competencies and intercultural competencies/ diversity awareness in the programme. Self-management competencies were mentioned particularly often, relationship skills were also relatively frequent. Many of the student groups further mentioned intercultural competencies/diversity awareness, but this was seldom in the interviews with school staff. Self-awareness was mentioned often by school leaders/other school staff and also by some teacher and student groups. Social awareness was rather seldom and responsible decision-making was hardly mentioned by any of the stakeholders. Considering that inner exercises made up a large part of the programme, it is apparent that the programme indeed had a strong focus on self-management and self-awareness. Many of the activities and games as well as exercises involving discussion and dialogue further addressed relationship skills, and the student programme contained more exercises for increasing intercultural competencies/diversity awareness than the school staff programmes. Hence, the experiences of participants' largely reflect the programme priorities. Apart from supporting the development of participants' Social and emotional competencies and intercultural competencies/ diversity awareness, the programme also had the aim to change school or classroom climates, and such changes were also mentioned by participants in response to the questions about what they would highlight as perceived outcomes of the training.

Altogether, these results suggest that the programme was helpful for developing SEI-competencies and improving classroom climates. It should, however, be noted that many participants said they "learned how to..." For example, one teacher group said they learned "*how to approach professionally conflict situations that occur regularly in schools*" and one student group said they learned "*how to better communicate with others*". Thus, the participants acquired knowledge about strategies for handling social situations. It remains open, however, whether they are also able to and willing to use this knowledge in everyday situations. Further several participants said that they learned social norms such as "*discrimination is a bad thing*". This also does not necessarily mean that they will in the future actively try to reduce discrimination in school. Hence, even though most of the interview groups could list many things that they thought they had learned through the programme, it remains open whether this learning actually has had an effect on participants' behaviours and school

practices. In any case, knowing how to do something is a necessary precondition for doing it. So even if participants only learned the former and not the latter, it can still be considered a relevant learning effect, just one that would need to be further developed in the future to achieve the aims of the HAND in HAND programme.

Unintended effects were not mentioned during any of the focus-group interviews. Some participants used more general categories for describing the programmes' outcomes and did not refer to specific competencies. For example, some teachers said they felt empowered or that they learned new exercises for use in the classroom. Yet, all of the "other" outcomes described by participants were related to the intended outcomes. Hence, participants' perception of what they have learned through the HAND in HAND programme is largely in accordance with the programme's aims and the participants noticed no negative effects from the programme.

8.4.2. How did participants like the programme?

Participants' evaluations of the programme were largely positive. Many participants liked and enjoyed the programme. Some found it "*fun*", "*interesting*", "*innovative*", "*useful*", "*helpful*", "*successful*", or even "*an excellent experience*" or "*the best training so far*". Yet, stakeholders differed considerably in their evaluations: Most groups of teachers as well as groups of school leaders and other school staff evaluated the programme positively. In contrast, there were quite a few student groups that gave the programme a mixed evaluation and one student group (out of 18) even explicitly disliked the programme.

Why are the evaluations of adult participants more positive than those of the students? Some responses to the interview questions suggest that teachers found the training particularly helpful, because it addressed common problems and needs of teachers arising from the complexity and contrariety of the demands of their profession. In particular, learning to better cope with stress and handle conflicts with students was considered valuable by many teachers. Accordingly, some teachers said they felt "*empowered*" by the training and others noted that they felt appreciated and respected because the training was about them and their own well-being, not only about student learning. The school leader/ other school staff training was considerably shorter than the teacher training and this was actually criticized by these stakeholders. Nevertheless, they also evaluated the programme quite positively.

Unlike the adults, students could apparently not be reached so well by the HAND in HAND programme. Several reasons are conceivable: First of all, it might be that the exercises were not well adapted to the needs of 13/14 year olds. In fact, some student groups said that they found the exercises too “childish” or that they already knew many of the things taught in the programme. There was, however, a significant overlap between the exercises the students were doing and those that the adults were doing in their trainings, and none of the adults complained that the content of the training was not serious enough for them. The games and physical exercises are indeed playful. Possibly, 13/14 year olds have a strong desire to distance themselves from the childhood they have just left behind, so that playful activities are appreciated less by this age group than by other age groups (including adults). Another possible explanation is that it was less obvious for students how learning self-management and relationship skills can bring an improvement for them personally – in particular, considering that the effects of inner exercises often do not become immediately visible but rather require some effort and regular practice. Students might have different experiences of stress and conflicts in school than the adults and, consequently, their motivation to commit themselves to the training might have been lower from the beginning. Thirdly, students at the age of 13 or 14 years might be sceptical of adults' proposals more generally, especially in schools which they do not attend out of their own free will. Accordingly, a meta-analysis by Yeager, Fong, Lee and Espelage (2015) suggested that the existing anti-bullying programmes are much less effective with students in grade eight or older than with younger students. Referring to this study, Downes and Cefai (2016) also stated: “With older students, the question also arises as to their particular resistance to didactic style approaches that would undermine their increased sense of autonomy” (p. 39). In this context, it might have also played a role that students were not involved in the decision to participate in the HAND in HAND programme (as we also know from the interviews). According to the responses of teachers and school leaders/ other school staff to the question, why the school decided to participate in the HAND in HAND programme and who decided, students had not been involved in any of the participating schools (on a side note: in more than half of the schools teachers were also not involved in this decision). This might have caused some scepticism and defence on the side of the students from the beginning. Still, there were also many students who liked the HAND in HAND programme.

8.4.3. Which criteria did participants use to evaluate the HAND in HAND programme?

Participants used a variety of different criteria for their individual evaluations of the HAND in HAND programme. However, unlike the developers of the programme, hardly any participant referred to the programme's effectiveness for explaining how they came to their assessment of the programme's quality. A few interview groups did mention changes to the school or classroom climate but only a few, and hardly any group argued that they liked the training, because it had fostered their Social and emotional competencies and intercultural competencies/diversity awareness. Hence, participants' evaluations were based on criteria different from those used in the experimental evaluation of the HAND in HAND programme.

Participants rather mentioned short-term than long-term changes as criteria: Many of the groups liked the positive atmosphere during the training sessions and that the programme had made the participants feel good. Considering that two central aims of the HAND in HAND programme were to foster the participants' emotion regulation and to improve school climates, these short-term effects on participants' moods and the group atmosphere can actually be considered one step in the right direction. It could, therefore, be interesting, if future evaluations of the HAND in HAND or similar programmes would explicitly assess such short-term changes. Apart from these short-term changes, most evaluations of participants were, however, based on criteria unrelated to effectiveness.

The most important criterion underlying students' assessments of the quality of the programme was whether the programme was "*fun*" or "*boring*". These words were used quite often by students in response to the question how they liked the programme overall. Teachers as well as school leaders and other school staff often mentioned positively the programme's focus on personal development and "*empowerment*" of teachers and on social relations. They indicated that they found these topics quite important. Also many adult participants liked the experiential approach of the training. Teachers further appreciated that they had gotten to know different types of exercises with different purposes (e.g. for energizing and for calming down a group) that they can use in the classroom. When adults gave the programme a more critical evaluation, then it was often because they found the programme too short, because the school staff specialized on social and emotional learning (school social workers, pedagogues and psychologists) had not been sufficiently included, or because they found it difficult to implement the exercises learned in the classroom.

8.5. Conclusion

The HAND in HAND programme received positive evaluations by most of the school staff. Many teachers, school leaders and other school staff liked the HAND in HAND programme. Students' evaluations of the programme were rather mixed: Some students liked the programme, others found it only ok. There were, however, very few students and no adults who explicitly disliked the programme. When participants liked the training, this was often because they found it "fun" or because it made them feel good and created a positive atmosphere in the group. Teachers further liked that they learned concrete techniques for their individual stress-management as well as exercises they can implement in the classroom. Many found this useful. It is less clear, however, how much of this participants have actually been able to implement. Only very few of the participants explained a positive evaluation of the HAND in HAND programme with its effectiveness. Nonetheless, they listed many competencies they think they have acquired during the programme, when they were asked directly what they think they have learned from the training. Yet this does not necessarily mean that they are able or willing to use their knowledge to actually manage their behaviour in everyday social interactions. There were only few participants who explicitly mentioned changes of their own behaviour or school practices as a result of the programme. So findings from the interviews suggest that participants have gone the first step for improving their SEI skills, i.e. learning how to address and cope with difficult personal and/or social situations, but from the interview responses it remains open, whether they have been able to use this knowledge in everyday life or not.

8.6. References

- Downes P. & Cefai, C. (2016). *How to Prevent and Tackle Bullying and School Violence: Evidence and Practices for Strategies for Inclusive and Safe Schools, NESET II report*. Luxembourg: Publications Office of the European Union.
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: Sage.
- Yeager, D. S., Fong, C. J. M., Lee, H. Y., Espelage, D. L. (2015). Declines in efficacy of anti-bullying programmes among older adolescents: Theory and a three-level meta-analysis', *Journal of Applied Developmental Psychology*, 37(1), 36-51.

PART C: RESULTS OF FORMATIVE EVALUATION

Chapter 9:
Participants' ideas for improving the HAND in HAND
programmes²²

Svenja Vieluf, Mojca Rožman, Nina Roczen

²² Please note that part of the analysis presented in this chapter has already been published in Vieluf, Denk, Rožman and Roczen (2020).

9.1. Introduction

Whereas the previous three chapters have all contributed to a summative evaluation of the quality of the HAND in HAND programme, the present chapter has a different focus. It aims at deriving suggestions for an improvement of the HAND in HAND programme. Hence, its purpose is formative, not summative.

Basis for this chapter is participants' feedback on the programme expressed during semi-structured focus-group interviews (see also Chapter 3). In particular, the chapter summarizes participants' responses to questions about their own concrete ideas for an improvement of the programme, but it also takes all other suggestions made during the interviews for an improvement of the HAND in HAND programme into account. Analysis method was qualitative content analysis (see e.g., Schreier, 2012). The categories were developed inductively (see also Chapter 3). First order categories concerned the overall focus of suggestions for improvement made during the semi-structured focus-group interviews: a) the overall focus of the HAND in HAND programme, b) the design of the HAND in HAND programme, c) the HAND in HAND exercises, d) the methods used by the trainers, e) programme implementation, e) support at the policy level, and f) organizational issues. Second order categories concerned the specific types of suggestions within a) to f).

9.2. Results of semi-structured focus-group interviews

Results of inductive content analysis of the semi-structured interviews with participants of the HAND in HAND programmes are presented separately for the three stakeholder-groups, students, teachers, and school leaders/ other school staff, in the following. The chapter ends with a conclusion, in which the three perspectives are summarized to derive concrete suggestions for an improvement of the HAND in HAND programme.

9.3. The perspective of students

Students commented on the overall focus of the HAND in HAND programme, the design of the HAND in HAND programme, the HAND in HAND exercises, the methods used by the trainers, programme implementation, support at the policy level, and organizational issues. However, comments concerning the overall focus of the HAND in HAND programme as well as comments concerning support at the policy level and organizational issues were rather rare. Often students commented on specific exercises.

9.3.1. The design of the HAND in HAND programme

Students commented neither positively nor negatively on the design of the programme when they were asked how they evaluated the programme, what did they like and what did they find challenging about the programme. However, six student groups (out of 18, i.e. 33%) suggested changes to the design when they were asked, how the programme could be improved: Five of these groups suggested to extend the programme over a longer period of time and one of these five groups additionally suggested doing the training together with students from other classes. The sixth group said that they would have liked to have a HAND in HAND training once a week over a whole day.

9.3.2. The overall focus of the HAND in HAND programme

Students mentioned the programme's focus (i.e. social, emotional and intercultural learning) only seldom during the semi-structured focus-group interviews. However, there were four groups who fundamentally criticised the focus (out of 18, i.e. 17%). Two of the three groups said that they already knew many of the things that were taught during the programme. Also two groups found the exercises "*childish*". One of the groups suggested to talk more about topics that are interesting for teenagers, "*e.g. relationships (boys and girls), new sports, addictions (drugs), sex*". Another group demanded "*a higher degree of seriousness that reflects our ability to reflect on more challenging topics*".

9.3.3. The HAND in HAND exercises

Many comments and suggestions of students made during the semi-structured focus-group interviews concerned the HAND in HAND exercises²³:

- Students were quite divided over the inner exercises. This type of exercises was often mentioned by student groups in response to the question whether there had been activities or topics during the programme that they were uncomfortable with or which they did not like so much. Students in four groups (out of 18, i.e. 22%) agreed on a negative evaluation of inner exercises and in six groups there was disagreement during the interview: some students stated they liked the inner exercises, others said they did not (out of 18, i.e. 33%). Hence, altogether, inner exercises were disliked by some (but not necessarily all) students in 10 out of 18 groups (i.e. 57%). However in all of the other eight groups (out of 18 i.e., 43%) the students said that they particularly enjoyed these exercises. Accordingly, five interview-groups (out of 18, i.e. 28%) suggested to have less inner exercises

²³ The types of exercises used in the HAND in HAND programme are described in Chapter 1.

when they were asked how the HAND in HAND programme could be improved, whereas two groups (out of 18, i.e. 11%) suggested including more of these.

- Physical exercises and games received critical comments from several groups (5 out of 18, i.e. 28%) and were suggested to be dropped by two groups (out of 18, i.e. 11%). At the same time, physical exercises and games were mentioned by nine groups (out of 18, i.e. 50%) in response to the question, what type of activities they had enjoyed most. Touching each other, on the other hand, felt uncomfortable for some (mentioned by one group out of 18, i.e. 6%). And the exercise “climbing up on a chair” was considered dangerous by another group (out of 18, i.e. 6%) where a child had fallen off the chair during this exercise.
- Exercises involving discussion or dialogue were particularly enjoyed by seven student groups (out of 18, i.e. 39%). Two student groups (out of 18, i.e. 16%) suggested having more of this type of exercises. On the other hand, four groups (out of 18, i.e. 22%) mentioned exercises with discussion or dialogue in response to the question of what they had felt uncomfortable with or disliked during the programme. Some students in these groups said they were not comfortable telling their classmates private things (mentioned in two groups; i.e. 11%).
- Exercises addressing diversity were liked by many groups: 13 groups (out of 18, i.e. 72%) mentioned them in response to the question, what types of activities they had enjoyed most in the HAND in HAND programme. Two groups (out of 18, i.e. 11%) suggested including more exercises addressing diversity in response to the question whether they had any proposals how the HAND in HAND programme could be improved. Only two groups (out of 18, i.e. 11%) mentioned them in response to the question of what they felt uncomfortable with or disliked, and these groups also were not against this kind of exercises: One student said that these exercises were “*not easy*”. Another student was irritated by how badly he*she had been treated when given the label “Roma” in a game.
- Other exercises: One student group suggested watching more films during the programme. Another group said they would like to do “*some role acting in a prescribed format*”. A third group suggested: “*a higher degree of seriousness that reflects our ability to reflect on more challenging topics*”.

9.3.4. Application of the new competencies outside of the training

Students in 6 interview groups (out of 18, i.e. 33%) said that they had done inner exercises after the workshop. One student said, he*she does it regularly every other day. Another student said he*she uses it as sleeping aid. In three groups were further students who used inner exercises to calm down when they were nervous before a test or to better concentrate when learning at home. In two groups the students who said they practiced inner exercises did not specify how often and with which purpose. So, there were some students who had used inner exercises for stress management outside of the training, but it was only single students and not in all schools (even though it should be noted here that we did not interview all students, so it is possible that there were also students at the other schools who had practiced inner exercises who were not interviewed).

9.3.5. The methods used by the trainers

Students also commented on the methods used in the programme: Work in pairs or small groups was seen critically by some students. Eight interview-groups (out of 18, i.e. 44%) expressed criticism with regard to working in pairs or small groups during the training. Two groups said that they had to work in groups who were not motivated to do the task and that they did not know how to react to that situation and found it difficult. One group problematized disagreements and conflicts during small-group work. Another group found it uncomfortable to talk to people with whom they did not get along well. Again another group said that they “*know everything about each other, so it was not interesting to talk again*”. Two groups found it uncomfortable to do exercises, in particular exercises that involved touching each other, in mixed gender pairs. In response to the question about proposals for improving the training two groups said that they would have liked to have less exercises in pairs.

9.3.6. Implementation of the programme

With regard to the quality of implementation of the programme, classroom management was mentioned by some student groups: Four groups (out of 18, i.e. 22%) said that they had disliked that other students were making noise and disturbing the programme. Students in one of these groups suggested that the trainers “*should have been stricter with discipline problems*”. Another issue that was mentioned (though only by one group out of 18, i.e. 6%) was that it felt uncomfortable when students who did not want to participate in exercises (which was explicitly allowed) were watching them doing the exercises. A different group (out of 18, i.e. 6%) criticized that trainers had used an academic language during the workshop. Finally, one suggestion from students that concerned

programme implementation was that students could have led some of the exercises (1 group suggested this out of 18, i.e. 6%).

9.3.7. Support at the policy level

Support at the policy level was not mentioned by students.

9.3.8. Organizational issues

One group (out of 18, i.e., 6%) suggested going out together after the workshop.

9.4. The perspective of teachers

Teachers also made suggestions for improvement concerning all the first order categories: the overall focus of the HAND in HAND programme, the design of the HAND in HAND programme, the HAND in HAND exercises, the methods used by the trainers, programme implementation, support at the policy level, and organizational issues. Particularly frequent were comments and suggestions concerning single exercises. Comments concerning policy support and organizational issues were rather rare.

9.4.1. The design of the HAND in HAND programme

The design of the programme was not mentioned in response to the questions what teachers had liked or what they felt uncomfortable with. However, several teacher groups suggested changes to the design of the programme: Most of these groups (9 groups out of 18, i.e. 50%) argued in favour of an extension of the programme, i.e. regular meetings over a longer period of time (e.g. once per month for a whole school year, or refreshment meetings once per year). Three teacher groups (out of 18, i.e. 17%) further suggested allowing participation of more teachers from their school (and possibly also from other schools). Also three groups (out of 18, i.e. 17%) were in favour of having a joint programme with students and/or other staff, at least for single training sessions. One group (out of 18, i.e. 6%) suggested involving more classes within their school. Four teacher groups (out of 18, i.e. 22%) argued that it would be better to work with younger students, i.e., with seventh graders or even with first graders.

9.4.2. The overall focus of the HAND in HAND programme

Asked, how they evaluated the HAND in HAND programme overall, what they liked in particular, what they found challenging, and whether they had any suggestions for improvement, teachers often said that they liked the focus of the

HAND in HAND programme, i.e. social, emotional and intercultural learning/diversity awareness. This pertained to 11 out of 18 teacher groups (i.e. 61%). Several teacher groups noted that they found the focus on social, emotional and intercultural learning/diversity awareness unusual; that the teacher trainings they had known often rather focused on how to foster student learning (7 out of 18 groups mentioned this, i.e. 39%). In particular, teachers in Croatia mentioned this. For instance, one teacher stated that it had been the first training that was “*focused primarily on the empowerment of the teachers*”. Another teacher said: “*It really seemed important to me that the focus was also on the teacher*”. It is noticeable that something about the training made some teachers feel recognised in a way that was apparently exceptional, as best illustrated by the following quote: “*That was what I liked about this programme: One felt important*”.

9.4.3. *The HAND in HAND exercises*

Many comments and suggestions of teachers concerned the HAND in HAND exercises. Half the teacher groups (9 out of 18, i.e. 50%) said that the exercises were useful. For example, one teacher said: “*it was applicable in the classroom, but we also learnt the techniques for self-awareness and personal growth which we as teachers need as it is a stressful job*”. Another teacher stated: “*There were new exercises that felt applicable to the school and that can be used in everyday life*”. In addition, 3 groups (out of 18, i.e., 17%) emphasised that they had become familiar with a variety of exercises. An example statement is: “*It was a good mix of exercises, group strengthening, physical, relaxation. A good package*”. Three groups (out of 18, i.e. 17%) positively viewed the training for not being presented as a fixed sequence of exercises, but as a fund of ideas, exercises and techniques from which they could choose whichever seemed most suitable for their situation and purpose. Two groups (out of 18, i.e. 11%) said that they would have liked to get to know even more exercises that can be used in the classroom. Two groups (out of 18, i.e. 11%) criticized that the exercises were repeated in different training sessions. Contrariwise, another group (out of 18, i.e. 6%) said that they would have liked to have more repetitions of exercises. Hence, teachers liked that the programme contained different types of exercises and found these exercises generally useful. Still, not every teacher liked every exercise and most groups made more specific comments on different exercise type, which are summarized in the following:

- The inner exercises were liked by many, but not by all teachers: 12 teacher groups (out of 18, i.e. 67%) mentioned this exercise-type in response to the question about what had worked well for them in the training and one group

(out of 18, i.e. 6%) suggested having more inner exercises in the programme. At the same time, six groups (out of 18, i.e. 33%) mentioned inner exercises in response to the question of what the teachers had experienced to be challenging in the training. Most of the latter groups did not express dislike for the inner exercises, but rather found them unfamiliar and difficult in the beginning. Nevertheless, two teacher groups (out of 18, i.e. 11%) also suggested having less inner exercises in the training.

- The physical exercises and games received more mixed evaluations from teachers. They were mentioned by five interview groups (out of 18, i.e. 28%) in response to the question what they had liked about the training. Accordingly, in one of the teacher groups (out of 18, i.e. 6%) it was suggested to include more physical exercises in the HAND in HAND programme for teachers. On the other hand, five interview groups said that they found physical exercises challenging (out of 18, i.e. 28%). One of the latter groups explained that they had been uncomfortable with being touched by people they had not known well. Another group found the chair exercise (climbing up and down on a chair) dangerous.
- When teachers identified specific exercises as challenging, these were often exercises involving discussions or dialogue – especially in Slovenia (altogether 8 out of 18 groups mentioned this, i.e. 44%). Four of the eight groups who said something critical about this type of exercise also suggested during the interview to reduce the time used for dialogue exercises during the programme. Reasons given for the critical evaluation of this exercise were that teachers had felt uncomfortable opening up in front of their colleagues, that awkward questions were asked during the dialogue exercise, and that teachers had been disappointed that no solutions were developed. Nevertheless, two teacher groups (out of 18, i.e. 11%, both Croatian) also said that they had enjoyed the dialogue exercises in particular.
- Exercises addressing diversity were liked in particular by three teacher groups (out of 18, i.e. 17%) and none of the interview-groups mentioned them in response to the question about things that had felt uncomfortable.

One group (out of 18, i.e., 6%), however, criticized that they felt “a lack of 'intercultural' in this project”.²⁴

9.4.4. Application of the new competencies outside of the training

Did teachers use the specific skills learned during the programme in everyday life? In 12 groups (out of 18, i.e. 67%), single teachers said they had tried to practice inner exercises outside of the training. Moreover, there were hardly any teachers who said they practiced regularly. 1 teacher said he*she practiced every day and a second teacher that he*she practiced once per week. The responses during 3 interviews were not clear with regard to the frequency of practice. In all the other groups (7 groups) the responses suggest that teachers practiced only occasionally.

Did the teachers implement the exercises learned during the programme in the classroom? In 13 teacher groups (out of 18, i.e. 72%) were teachers who said that they had tried activities and exercises from the programme in the target class. In 11 groups (out of 18, i.e. 61%) were teachers who had tried physical exercises or games, in 9 groups (out of 18, i.e. 50%) were teachers who had tried inner exercises with students, in 6 groups (out of 18, i.e. 33%) were teachers who used exercises addressing diversity with the target class, and in 4 groups (out of 18, i.e. 22%) were teachers who had used exercises involving discussion or dialogue. Hence, teachers did use the different exercises they learned in the classroom, but not all teachers implemented all types of exercises. Also, most of them used the exercises only occasionally (this pertains to 10 of the 12 groups who had reported to use them at all). Moreover, only 5 student groups (out of the 18 whose teachers had participated in the teacher training, i.e. 28%) noticed that their teachers had used exercises from the training during lessons in their class.

During the interviews 16 teacher groups (out of 18, i.e., 89%) further addressed difficulties with using the exercises in the classroom. 10 groups (out of 18, i.e., 56%) said that there was not enough time for including the exercises in their lessons; that they needed all time to push through the curriculum. Accordingly, one teacher group (out of 18, i.e. 6%) suggested introducing an extra lesson that is fully dedicated to social, emotional, and intercultural learning and another (out of 18, i.e. 6%) suggested including SEI in the subject curricula. Apart from a lack

²⁴ HAND in HAND exercises came more from the tradition of diversity awareness trainings than from that of intercultural trainings, i.e., they often included reflections of group processes and power structures and did explicitly not thematise differences between cultures.

of time also a lack of space and a lack of materials were brought up. Three groups (out of 18, i.e., 17%) said that they had no adequate space in the school for such kinds of activities, two groups (out of 18, i.e. 11%) problematized the class sizes, and 1 group (out of 18, i.e. 6%) brought up the costs for materials that they had to buy themselves. But not only external causes were mentioned. In three teacher groups (out of 18, i.e. 17%) the teachers said during the semi-structured focus-group interviews that they had forgotten the exercises after a while. Accordingly, 2 teacher groups (out of 18, i.e. 11%) said that it would have helped them to get handbooks and videos and another group said that regular reminders would have been beneficial for them. 2 teacher (out of 18, i.e. 11%) groups further mentioned that they felt insecure or uncomfortable doing these exercises with students. 1 teacher group (out of 18, i.e. 6%) was afraid that things would get out of control and two groups (out of 18, i.e. 11%) considered a poor climate in the classroom an argument not to implement the exercises in the class. In 1 group (out of 18, i.e. 6%) the teachers said that students had made fun of the exercises when they had tried them out. Further, 3 groups (out of 18, i.e. 17%) brought up that they had the impression that students felt embarrassed doing the exercises and 2 groups (out of 18, i.e. 11%) described a lack of motivation on the side of the students. 1 group (out of 18, i.e., 6%) said that students did not generally dislike the exercises, but got quickly bored by them, so that they cannot be used often in the same class.

In sum, many teachers found it difficult to implement the HAND in HAND activities in the classroom. Accordingly, a number of teacher groups said that they would have liked to get more support for implementation from their own school: One group (out of 18, i.e., 6%) asked for more support from the school leadership and more profound structural changes to enable implementation. One group (out of 18, i.e. 6%) said that implementation would have been easier if the subject curricula were less extensive. One group (out of 18, i.e. 6%) suggested developing school wide routines. Two groups (out of 18, i.e. 11%) said that it could have been helpful to organize the implementation of the HAND in HAND programme with the 'student health teams'. Other teachers would have liked to get more support for implementation from the trainers: Three groups (out of 18, i.e. 17%) would have liked to get a handbook and/or more specific guidelines on when and how to implement the exercises. Also three groups (out of 18, i.e. 17%) said that it would have been helpful to watch trainers do the exercises with students before trying to lead them by themselves. Again three groups (out of 18, i.e. 17%) said that it would have been helpful to get a reminder in between the trainings (e.g. via email). Finally, one group (out of 18, i.e. 6%) suggested getting more guidance throughout the process (regular supervision meetings) and another group (out of

18, i.e. 6%) suggested having more chances to practice during the trainings and over a longer period of time.

9.4.5. The methods used by the trainers

Nine teacher groups (out of 18, i.e. 50%) liked that the training had a hands-on approach. None of the teacher groups criticized or suggested changes to the methods used by the trainers.

9.4.6. Programme implementation

Seven teacher groups (out of 18, i.e. 39%) appreciated the positive atmosphere during the training. For example, one teacher said: *“The atmosphere was relaxed, we were very relaxed at the training too”*. Another stated: *“It was a good feeling and an atmosphere of acceptance”*. However, also some minor criticism of programme implementation was expressed: Two groups (out of 18, i.e. 11%) suggested better explaining the purpose of exercises before doing them.

9.4.7. Support at the policy level

Only one teacher group explicitly demanded more policy support. In this group teachers said: *“You need backing up from the top in the organization otherwise it will be up to the teachers.”* And: *“When you want change, you need help from the organization, the structures, then the teachers can plan the exercises themselves.”* Other teachers raised the topic of policy support rather indirectly: In 10 groups (out of 18, i.e., 56%) teachers said that they lacked time for implementing HAND in HAND in the classroom or that this conflicted with the curricula. The most obvious solutions to this problem would be including social, emotional and intercultural learning in the curricula (and removing other content accordingly) or introducing an extra lesson for social, emotional and intercultural learning. Hence, they indirectly suggested more policy support with this comment. Further policy support desired by teachers was providing materials (mentioned by 1 group out of 18, i.e. 6%) and adequate space/rooms (mentioned by 3 groups out of 18, i.e. 17%). Two groups (out of 18, i.e. 11%) problematized the class sizes (see also section d)). Finally, three teacher groups (out of 18, i.e. 17%) thematised the extra workload and how it was exhausting to participate in the training in the afternoon after a long working day, which indirectly suggests receiving time compensation for participation in activities like HAND in HAND – which would also be a form of policy support.

9.4.8. Organizational issues

Only single teacher groups commented on organizational issues: Three groups (out of 18, i.e. 17%) liked staying in a hotel. One group (out of 18, i.e. 6%) would have liked to have the training closer by (less time for travel). One group (out of 18, i.e. 6%) suggested doing the trainings with smaller groups. Another group (out of 18, i.e., 6%) said that it would have been helpful to make all materials available directly after the first session.

9.5. The perspective of school leaders and other school staff

Similar to students and teachers, also school leaders and other school staff commented on or made suggestions for improvement concerning all the first order categories, i.e., the overall focus of the HAND in HAND programme, the design of the HAND in HAND programme, the HAND in HAND exercises, the methods used by the trainers, programme implementation, support at the policy level, and organizational issues. School leaders commented particularly often on the design of the HAND in HAND programme. They also commented often on specific exercises, but not as often as students and teachers – which is probably due to the shortness of their training. Similar to students and teachers, school leaders and other school staff did not often comment on organizational issues or support at the policy level. Comments on or suggestions concerning the methods used and programme implementation from school leaders/ other school staff were also rather rare.

9.5.1. The design of the HAND in HAND programme

The design of the programme was often criticized by school leaders/ other staff. The main criticism voiced was that their own training had been too short (only 2 days): six groups mentioned this (out of 16, i.e. 38%). For example, one group said the training had been “*an initial spark, but it remained somehow unfinished*”. Another group expressed quite sharp criticism related to this, saying: “*I trust less and less in these isolated inputs. Similar results would be achieved if one would hang out with kids, do something with them, but not just workshops, projects...*” Half of the school leader/other school staff groups suggested extending the programme (8 out of 16, i.e. 50%). It was also suggested to bring more stakeholders into the training: One group (out of 16, i.e. 6%) said that it would be good to allow for participation of all teachers within a school who wanted to participate. Another group (out of 16, i.e. 6%) said that the training should address more students within each school. A stronger involvement of parents was also suggested by two groups. One of these two groups argued that a meeting of the project team with the parents could have improved acceptance of the programme

by students. Finally, two school leaders/other school staff groups (out of 16, i.e. 13%) argued that it would have been better to work with younger students.

One school leader/ other school staff group (out of 16, i.e. 16%) mentioned positively that HAND in HAND addressed different stakeholder-types (students, teachers and school leaders/other school staff). However, three groups (out of 16, i.e. 19%) criticized that these stakeholder-types were separated during the whole training, and one of these three groups also suggested bringing all three stakeholder-types together at for one or more training-session. The student health teams, i.e. school psychologists, social workers, etc. should have further been given a more active role in the process, said three groups (out of 16, i.e. 19%). In one of these groups the student health team and teachers for special needs said they had felt left aside. They suggested: “*We could be the motors instead of testing without practising before*”.

Something that school leaders liked about the design of the HAND in HAND programme was the European dimension of the HAND in HAND project (mentioned by 1 group out of 16, i.e. 6%). However, another group also mentioned critically that they would have appreciated getting the opportunity to exchange experiences with schools in other European school systems.²⁵

9.5.2. The overall focus of the HAND in HAND programme

Similar to the teachers, many school leader/other school staff groups generally liked the focus of the trainings. Five school leader/other school staff groups (out of 16, i.e. 31%) commented positively on the programme's focus on personal growth and empowerment and three groups commented positively on the programme's focus on relationship-building (out of 16, i.e. 19%). Four of these groups (out of 16, i.e. 25%, all 4 in Croatia) described the thematic foci as being “*new*” and “*different*” from other trainings. However, two groups (out of 16, i.e. 13%) suggested a slight change to the focus: They would have liked to talk in particular about children with special needs.

9.5.3. The HAND in HAND exercises

School leaders/ other school staff commented more seldom on specific exercises than the other two stakeholder-types. This might be due to the fact that they had

²⁵ Please note that such a chance for trans-national networking was given after the interviews had taken place at the final HAND in HAND conference, where teachers from all participating school systems were invited to participate.

received a considerably shorter training compared to teachers and students (only two days). The comments they made are summarized in the following:

- Inner exercises were liked in particular by five school-leader/other staff groups (out of 16, i.e. 25%, 2 Slovenian, 3 Croatian). One group (out of 16, i.e., 6%) suggested having even more of these in the programme. At the same time, four school leader/other school staff groups found the inner exercises difficult (out of 16, i.e. 25%). However, only one of the latter groups explicitly disliked them; the others found them only unusual and/or somewhat difficult.
- Two school leader/other school staff groups liked the physical exercises and games in particular (out of 16 groups, i.e. 13%), one group found these challenging (out of 16, i.e. 6%).
- Exercises involving discussion or dialogue were explicitly liked by two school leader/ other school staff groups (out of 16, i.e. 13%), but perceived as challenging by three groups (out of 16 groups, i.e. 19%). These latter groups said they found it difficult to listen without giving advice or that they had, more generally, not completely understood the exercise. One school leader/other school staff group said they found it difficult to open up in front of people they had not known before.
- Exercises addressing diversity were particularly liked by four school leader/other school staff groups (out of 16, i.e. 25%, all 4 Croatian) and criticized by none of the groups. Only one group said that they would have liked to have more exercises addressing intercultural competences, i.e. *“what to do with foreigners and Roma students, how to work with them in practice”*.

9.5.4. Application of the new competencies outside of the training

In 9 school leader/ other school staff groups (out of 16, i.e. 56%) were individuals who said they had done inner exercises outside of the training. However, it was only individuals, not all participants in these groups. Moreover, only few individuals said they did this on a regular basis. 4 school leader/other staff groups (out of 16, i.e. 25%) mentioned implementation in response to the question about challenges. Two groups (out of 16, i.e. 13%) mentioned a lack of time as a central barrier. One school leader said: *“I did not perform those exercises outside the training. To stick to this, during the daily routine, this is based on discipline, to exercise every day.”*

School leaders and other school staff were not explicitly asked whether they had used exercises in meetings with teachers or other school staff. However, two reported that they had tried and experienced difficulties. One said: *“When I came from the training, I was thinking that this is something that I would introduce at the staff meeting, but the group was too big and the space inappropriate”*. Another said: *“I was performing the relaxing techniques with half of the teachers three weeks ago. It was hard for them when they closed their eyes. Some of them went to the toilet at that time, because they did not want to do those relaxation techniques”*.

4 school leader/ other school staff groups (out of 16, i.e., 25%) suggested more support for implementation. One of them asked for specific implementation guidelines. Another suggested *“external control”*, i.e., to get homework assignments between training sessions. Similarly, the third group said *“the training could address problems one by one with a supervision session after each attempt at ‘problem solving’”*. Also a third group suggested *“supervision”* during and after the programme to support implementation.

9.5.5. The methods used by the trainers

Methods used in the programme were mentioned seldom: Three groups (out of 16, i.e. 19%) praised the experiential approach of the training.

9.5.6. Programme implementation

One group (out of 16, i.e. 6%) liked that there was no time-pressure during the training sessions; that everything was *“relaxed”*.

9.5.7. Support at the policy level

Policy support was only demanded by one of the 16 school leader/other staff groups (i.e. 6%). In this group one school leader said: *“In my opinion, all things would have to be implemented into the curriculum if we want to raise and form a healthy society.”*

9.5.8. Organizational issues

One group (out of 18, i.e. 6%) said that they liked that the programme did not hinder the school process because it took place outside of lesson time. Another group (out of 18, i.e. 6%) suggested getting handouts.

9.6. How can the HAND in HAND programme be improved? Integrating feedback and suggestions from the three stakeholder-types.

From an integration of the perspectives of different stakeholders, several suggestions for an improvement of the programme can be derived.

9.6.1. The design of the programme

The most frequent suggestion made by all stakeholders during the semi-structured focus-group interviews was extending the HAND in HAND programme over a longer period of time. School leaders and other school staff, who had only two days of training, said that they had merely been shown “*a glimpse into the programme*”, but also teachers and students – who had five whole training days and five modules á 90 minutes respectively – suggested having more training sessions. For example, single training days once per month during a whole school year have been suggested by participants. It is remarkable that many of the different stakeholders participating in the HAND in HAND programme felt that a longer training would have been beneficial.

Merely adding training-sessions with more exercises would probably not enhance training effects substantially, however. Rather, the interviews suggest that it would be helpful to complement the existing training with extra sessions that have the aim to support the application and implementation of competencies, knowledge and techniques acquired during the programme to everyday situations and, in particular, to classroom teaching, since many of the adult groups described this as challenging during the interviews, and only few actually implemented the techniques and exercises learned in the programme on a regular basis. What might such support for implementation look like in practice? Participants suggested that the trainers could provide guidelines and send regular reminders. Teachers further proposed supervision. Following these suggestions, additional training sessions with teachers could, first, be used for discussing ideas and plans for implementation, and then, later, for reflecting on teachers' experiences with implementation and difficulties that might have arisen. Possibly, a mobile phone app could also be offered for regular reminders to practice inner exercises to all those who are interested in getting support for their practice.

Apart from the time-related extension of the HAND in HAND programme, several participants also suggested addressing more students and more teachers within each school or even the whole school. Two school leader groups further argued that offering workshops or at least an informative meeting for parents

could improve students' acceptance of the programme. Moreover, several interview groups suggested better integrating the three trainings, i.e., that different stakeholders should have at least part of the training together. A related issue is better involving the specialized school staff in the planning and implementation of the programme. School social workers, school psychologists, school counsellors, special needs teachers, student health teams and similar professions are, among other things, also responsible for helping with conflict resolution in the school and with supporting social, emotional and intercultural learning. Hence, it would be consequential to regard them as motors of change for school development processes aimed at improving the quality of social processes in the school. This desire was expressed by other school staff themselves during the interviews, but similar suggestions were also made by school leaders and teachers.

One additional design-related suggestion that appeared in several interviews with teachers as well as with school leaders and other school staff was starting at an earlier age with the programme. The school staff felt that it would be easier to have an impact with younger students and that younger students might be more open and motivated to participate in the programme. They also argued that many of the things taught in the programme would help improving school processes and, hence, should be learned as early in the school career as possible. This impression from practice is in accordance with research results: For example, a meta-analysis by Yeager, Fong, Lee and Espelage (2015) suggested that anti-bullying programmes are much less effective with students in grade eight or older than with younger students.

9.6.2. The overall focus of the programme

The purpose and thematic focus of the training was well received by the adult participants, but less by the students. With regard to the school leaders this result is not surprising considering that it was them who decided that the school should participate in the HAND in HAND programme. The other staff and teachers – who also commented very positively on the focus – were not always but quite often involved in this decision as well (other school staff in 44% of the schools; teachers in 39% of the schools). The students, in contrast, who had the most critical attitude towards the theme and content of the training, had not been consulted with regard to the decision about participation in any of the participating schools. This could possibly explain the difference between stakeholder-types with regard to the overall attitude toward the training. It might be that adults find social, emotional and intercultural learning in schools generally more important than students do, or that the school staff programmes were better tailored to the

needs of adults than the student programme were tailored to the needs of adolescents. It should, however, also be considered that only schools self-selected for the HAND in HAND programmes where adults were already interested in these topics beforehand – it was the adults (the school leaders and often also other school staff and/or teachers) who decided that their school should participate. This self-selection process concerning participation in the HAND in HAND programmes did not take students interests and opinions into account – they were not asked in any of the participating school whether they supported participation in the HAND in HAND programmes. This also brings up another issue that might explain the difference between stakeholders in their evaluations of the programmes: Possibly, students would have experienced more ownership of the programme and more intrinsic motivation to participate had they been more involved in the decision that their school and class would participate. It might, thus, be an interesting experiment to repeat the HAND in HAND study, but allow for more participation of students in the decision whether a school should participate or not.

Better supporting students' autonomy and allowing for more participation – not only in the decision that the school would participate in the programme, but also regarding content and organizational issues – could be one strategy to increase students' enthusiasm for the HAND in HAND programme and of similar programme's more generally. For example, discussing results of a meta-analysis of anti-bullying programmes, Downes and Cefai (2016) stated: "With older students, the question also arises as to their particular resistance to didactic style approaches that would undermine their increased sense of autonomy" (p. 39). Hence, the programme itself could be modified to allow more participation in decision making and better support participants' sense of autonomy.

9.6.3. The HAND in HAND exercises

With regard to an evaluation of the specific exercises used during the HAND in HAND programme, the interviews did not provide a clear picture. Some participants liked some types of exercises better and others other types of exercises. Consequently, the mix of exercises that is realized in the HAND in HAND programme appears optimal for reaching as many participants as possible (some exercises will reach some participants and others other participants, so mixing them maximizes the number of participants reached – apart from the fact that different exercises had different purposes). The mix of different exercises was also explicitly appreciated by participants during the semi-structured focus-group

interviews, in particular by teacher groups. However, two types of exercises received less ambiguous feedback:

- First, student groups were fairly united in their positive evaluation of *exercises addressing diversity*. It seems that group processes, social inequalities and discrimination are topics that are interesting for students around 14 years (across different European school systems). This may be taken as indication to increase the relative share of this type of exercises in the student programme.
- Second, many of the teacher and school leader/ other staff groups had difficulties with *the dialogue exercise* and several groups suggested making them shorter or dropping them. Maybe their share should, however, not be reduced – as they have an important function in the programme. Yet, the critical feedback of participants suggests that it might be advisable to modify the dialogue exercises in a way that makes participants feel more confident and comfortable with them. Possibly, they need a more detailed introduction and more modelling by the trainers – considering that several groups said that they had not completely understood the exercise and found it difficult. Also, it might be good to practice feedback rules more explicitly before doing this task, as some teachers complained that the feedback they had received by colleagues was not always cautious and some even felt hurt. So including a feedback training might be helpful.

Another specific criticism that appeared during the interviews with different stakeholders was minor mishaps during the chair exercise, indicating that this exercise can involve a certain risk of injury and also that it is somewhat exclusive for participants who are (for any reason) less athletic. It might be advisable to drop this exercise from the training or include a specific warning that group leaders need to consider well how athletic the participants are before including it.

9.6.4. *The methods used by the trainers*

The methods used in the training were well-received by the adult participants and only few changes were suggested by these groups. In contrast, students were more critical of the methods. More specifically, they did not like the work in pairs or small groups. Some of them would have preferred to choose their working partner by themselves. They did not want to work together with students with whom they were not friends or with students having a gender different from their own. Some students further reported conflicts and being annoyed by fellow students who were distracted during group work. There are many theoretical and practical arguments in favour of small group work and also in favour of random allocation of students

to groups (Lotan, 2006), so refraining from small group work does not appear advisable – in spite of students' criticism. However, it might be helpful to include a skill-builder with the purpose of training those social skills needed for productive group-work at the beginning of the training in classes that are not used to cooperative learning (see e.g., Cohen & Lotan, 2014). Maybe students could further be allowed to choose their partners by themselves for specific exercises, namely for those that involve the exchange of very private information and/or touching each other – at least in the beginning of the training before a higher level of trust has been established in the classroom.

9.6.5. Programme implementation

Implementation of the programme was evaluated positively during most of the interviews. Many participants enjoyed the good atmosphere during the training and some even praised the trainers. However, again, students were more critical than adults. Several student groups complained about fellow students making noise and disturbing the training. They suggested stricter classroom management. Another student group argued that the disciplinary climate became better over the course of the training and assumed that there might be no problems with noise and disturbances after a while – if the training had taken place more often. It was already discussed above, that many participants were in favour of extending the training. Possibly, this would already help with disciplinary problems, as the students assumed. Whether this is the case or not, it might still be helpful if future trainings of trainers for the HAND in HAND programme would address this issue more extensively. Also, it could make it easier for trainers if the training would be modified to become a bit more supportive of students' autonomy. During the HAND in HAND trainings, students could choose not to participate in any exercise they did not like. Also, respecting students, taking serious what they think and say is an important principle of the training. Hence, the HAND in HAND programme realizes several criteria for being autonomy supportive. Yet, students were not involved in the decision to participate in the programme in the first place. Neither had they choices with regard to the activities and exercises or methods used during the programme sessions. Better supporting students' autonomy could increase their intrinsic motivation to participate in the programme (see e.g. Deci & Ryan, 1987; Ryan & Deci, 2000; for a further discussion of autonomy support strategies see e.g. Assor, Kaplan, & Roth, 2002; Deci & Ryan, 1987; Ryan & Deci, 2000; Stefanou, Perencevich, DiCintio, & Turner, 2004) and, thereby, contribute to a better disciplinary climate in the classroom.

Two smaller suggestions concerning implementation made by participants were: Explaining better the purpose of exercises and using a more simple language.

9.6.6. Support at the policy level

Explicit demands of policy support were seldom during the interviews, but implicitly quite a few teachers and some school leaders/ other staff addressed this topic. Some participants argued that the programme was too short to show any effects and many participants suggested an extension of the programme – extending the programme would require considerably more funding and, hence, more policy support. Second, some teachers mentioned the extra workload the programme entailed and also argued that they lacked the time to carefully plan, organize and reflect implementation in the classroom – time compensation for participation would also need to be decided at the policy level. Finally, single groups suggested including social emotional and trans-/intercultural learning in the curricula instead of supporting single workshops.

9.6.7. Organizational issues

Several small suggestions concerning the organization of the programme were made by single groups: First of all participants would have liked to get handouts. Second, they suggested doing the training with smaller groups. Third, one teacher group would have liked to have the training nearby so that they would have had to spend less time on travel.

9.7. Conclusions

In sum, participants made a variety of suggestions, how the HAND in HAND programmes could be further improved. Some topics were addressed by multiple groups and different stakeholders. In particular, extending the programme over a longer period of time and providing more support for implementation were recurring themes. It is further noteworthy that students' evaluations of the programme were significantly more critical than those of adults and that the lack of motivation on the side of the students also manifested itself in noise and disruptions during the training. At the same time, it is striking that students were the only stakeholders who were neither involved in the decision to participate in the programme nor in decisions about the focus of the programme, the choice of exercises or local adaptations. Possibly, allowing for more participation and better supporting the autonomy of the students during the training could increase their intrinsic motivation to participate.

9.8. References

- Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours in predicting student's engagement in school work. *British Journal of Educational Psychology*, 72, 261-278.
- Cohen, E. G., & Lotan, R. A. (2014). *Designing groupwork: Strategies for the heterogeneous classroom*. New York: Teachers College Press.
- Deci, E. L. & Ryan, R. M. (1987). The support of autonomy and the control of behaviour. *Journal of Personality and Social Psychology*, 56(6), 1024–1037.
- Downes P. & Cefai, C. (2016). *How to Prevent and Tackle Bullying and School Violence: Evidence and Practices for Strategies for Inclusive and Safe Schools, NESET II report*. Luxembourg: Publications Office of the European Union.
- Johnson, D. W., Johnson, R., & Holubec, E. (2013). *Cooperation in the classroom* (9th ed.). Edina, MN: Interaction Book Company.
- Lotan, R. A. (2006). Managing groupwork in the heterogeneous classroom. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 525–539). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ryan, R. & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 78–88. Doi: 10.1037/0003-066X.55.1.68
- Schreier, M. (2012). *Qualitative content analysis in practice*. London: Sage.
- Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychologist*, 39(2), 97-110.
- Yeager, D. S., Fong, C. J. M., Lee, H. Y., Espelage, D. L. (2015). Declines in efficacy of anti-bullying programmes among older adolescents: Theory and a three-level meta-analysis'. *Journal of Applied Developmental Psychology*, 37(1), 36-51.

Chapter 10:
Quality assurance in the HAND in HAND project

Maria Rasmusson, Magnus Oskarsson, Nina Eliasson, Helene
Dahlström

10.1. Introduction

The overall aim of the quality assurance has been to ensure that the HAND in HAND project has good quality when it comes to planning and goal setting, organisation and execution and the project's final deliverables. In this chapter the method as well as the results of this work are described.

10.2. Method

This section describes the methods used in the quality assurance of the HAND in HAND project in the different stages and steps. The quality assurance work started in the planning stage when the application was developed. In the project's work packages, clear and measurable goals about procedures, meetings, tasks and outputs were formulated. Effort was taken to make sure the goals and objectives were formulated in understandable terms and could be measured. When the project was approved, a more detailed plan for quality assurance was produced. It was decided to focus on three main project areas: planning and goal setting, organisation and execution, and the project's final deliverables. Indicators used to measure the quality were process, performance, and outcome indicators. The process indicators were defined as the level of implementation of the activities, their conformity with the project proposal's provisions, keeping up with the project time-frames and schedule, and the dissemination channels used. The performance indicators included the level of the team spirit and collaboration and the number of target-group representatives involved in activities. The outcome indicators included the type and content of the outcomes, the quality of the outputs, and the outcomes' conformity with the parameters stated in the proposal.

10.3. Procedures and instruments

In practice, a set of procedures and instruments was developed in order to ensure the quality of these three areas. The methods, procedures and instruments used in the quality assurance were: (1) a web-based checklist to keep track of progress, including all work packages in the project and all expected outcomes and activities in each work package. At the beginning of the project (2) a risk management strategy was developed, and all partners contributed with possible risks, the level of severity, and how to manage them. After each project meeting (3), questionnaires about the meeting were delivered to all participants to measure the meeting's quality.

10.3.1. Web-based checklist

All deliverables planned in the project were included in a web-based checklist, together with information on which partner(s) hold the main responsibility for each deliverable as well as the starting time and the deadline. All partners were given access to the web-based checklist and asked to indicate when they had completed their task. The deliverables in the checklist served as indicators for monitoring the project's overall progress as well as for each work package. This served as a tool for monitoring but also as a way for all the partners to gain an overview of the overall complexity of the project and to create a shared understanding of the tasks needing to be accomplished.

10.3.2. Risk management strategy

Another part of the quality assurance work has entailed developing a risk management strategy. The strategy aimed at finding serious risks and possible solutions in advance (Olsson & Skjöldebrand, 2008). At the start of the project, each partner defined major risks within their area of responsibility together with a suggestion on how to manage them. The project coordinator has been responsible for monitoring and taking appropriate actions to prevent risks identified as being highly probable and severe.

For each identified risk, we estimated the likelihood of its occurrence, the severity, and possible measures to prevent or handle it.

10.3.3. Quality visits

The quality team visited the partners in Slovenia and Croatia during the field trial and conducted interviews with the national team. One class in Slovenia was also visited during the student programme. Moreover, the quality team completed a self-report about the work in Sweden

10.3.4. Meeting questionnaires

A web-based questionnaire has been delivered after each project meeting. The planning of the meetings, preparations and decisions taken during the meeting are monitored. The scope of the quality assurance of these meetings has been to assure high quality communication within and among all partners during face-to-face meetings. The questionnaires were administered to all participants of each meeting. The quality assurance team analysed the data after each meeting and reported the results to the project manager, including suggestions for improvement if needed.

10.4. Results

In this section, the results are reported and organised according to the three main areas that were in the focus of the quality assurance: the project's planning and goal setting, organisation and execution, as well as the final deliverables.

10.4.1. Quality assurance of the project's planning and goal setting

In stage one, *planning*, the project proposal was developed by the Educational Research Institute (ERI) in Slovenia and reviewed by all of the partners. The application procedure had two steps whereby a shorter proposal was first submitted and reviewed by the Erasmus+ committee. The present project was chosen and invited for the second step. Thus, the full proposal was developed, submitted and approved. The project was not funded with the proposed amount, and the lower budget induced a review of the proposal and saw changes being made to the project plan.

The application of HAND in HAND consists all work-packages and all deliverables described in detail. Each partner has held distinct roles and responsibilities. The deliverables have been transformed to a web-based checklist in order to make them well known and transparent. All of the partners have recorded which deliverables are ongoing and when they are completed.

10.4.2. Risk management strategy

An essential part of the work on the quality assurance plan has been to develop a risk management strategy. At the beginning of the project, each partner defined major risks within their area of responsibility together with a suggestion for how to manage them. The project coordinator has been responsible for monitoring and taking appropriate actions to prevent risks identified as being highly probable and severe.

For each identified risk, we estimated the likelihood of its occurrence, the severity, and the prevention measures.

Examples of some severe risks:

- Time delay in programme/instrument development, sampling, data collection etc. which reduces the time left for analysis and report writing. To manage this risk, all partners need to meet the deadlines.

- Schools may drop out of the programme. This risk could be managed by sampling replacement schools.
- Fragmentation of actions across partners. Clear communication should reduce this risk.
- Low alignment between the content of the workshops and the assessment. This could be prevented by careful operationalisation of the goals and targets.
- No stable group of local trainers throughout the project. To prevent this, we need to urge the partner school systems to assemble a stable group (e.g. that the 2 persons who are going to be the teacher trainers also participate in all the training arranged by the responsible partner).
- Many levels of adaptation can affect the outcome measure. Keeping the focus on implementation of the core values in the project can decrease this risk.
- Too few schools willing to participate in randomised conditions. We could reduce this risk if the project is well communicated and the schools are contacted and prepared in time.

10.4.3. Quality assurance of the project's organisation and execution

This part describes the information and workflows, the quality of communication among partners, the partners' timeliness according to the project agenda, and partner satisfaction.

All partners have made a brief report to the ERI every month about progress, risks and drawbacks. The ERI has included these reports in the HAND in HAND monthly newsletter. Besides the project meetings, monthly online meetings for monitoring the project have been arranged by the managing team.

One measure was to evaluate the project meetings. Table 10.1. presents the results of questionnaires completed by all partners attending the project meetings.

Table 10.1.

Assessment of the first four project meetings (PM)

		How did you perceive the main purpose of the meeting? (Unclear 1- Clear 5)	I felt that we had a meeting characterized by the sensitivity, responsiveness and trust. (Not 1- Agree 5)	It was clear what the meeting decided (Unclear 1- Clear 3)	clear It is clear what the next step in the project is for me after the meeting (Unclear 1- Clear 3)	The timeline about what to do after the meeting is clear (Unclear 1-Clear 3)
PM1	13	4.5	4.6	2.3	2.8	2.8
PM2	14	4.5	4.6	2.9	2.9	3.0
PM3	16	4.6	3.9	2.8	3.0	2.9
PM4	20	4.6	4.6	2.9	2.9	3.0
PM 5	15	4.8	4.9	2.9	2.9	3.0

Table 10.1. reveals that the project partners are very satisfied with the meetings. All partners answered that they prepared their contributions to the meetings on time, they actively participated in the meetings and felt they had opportunities to ask questions and that these questions were discussed in the meetings.

10.4.4. Quality assurance visits to the partners

Two quality assurance visits were made to the project partners in Slovenia and to the partners in Croatia during the HAND in HAND field trials in November 2018. Interviews were performed with the team members at the project partner institutions during these visits. In Slovenia, it was also possible to visit a school where the student programme was being implemented.

a) Slovenia

Overall, HAND in HAND in Slovenia progressed according to the plan. The sampling procedure, the translations, contacts with the participating schools, organisation of the materials, and collecting parental consents were accomplished. The HAND in HAND field trial and implementation of the student programme went according to plan. There were some minor adaptations of the modules. All practices were implemented but the order of the practices in some cases was switched in order to meet the dynamic of the student group. Some exercises were also shortened when the students' attention started to wane. Some topics were switched when working with the teacher programme so as to fit in with the mood of the teachers. One conclusion is to be sensitive and be aware of the teachers'

mood when implementing the teacher programme (see e.g. Lund Nielsen et al., 2019).

Obstacles: A member of the Slovenian team went on sick leave and a new person had to assume their tasks, which was achieved. Early on, before the field trial started, two schools dropped out and two new schools had to be included. This was resolved by recruiting replacement schools that had been selected during sampling for that purpose.

b) Croatia

In Croatia, the visit showed the same results as in Slovenia; the project has to that point progressed according to plan; the sampling procedure, the translations, contacts with the participating schools, organisation of the materials, and collecting parental consents.

Obstacles: The process in the project was new to the Croatian team members and much effort was made to understand the organisation, the work packages, and all the tasks in the project. They also experienced some difficulties convincing schools to participate in the project.

c) Self-report from the Swedish team

The programme has progressed according to plan. All materials have been translated with the help of professional translators, with one team member being responsible for the necessary adaptation. Good contacts with regional and national stakeholders ensure the project is well known and this also helped when recruiting schools for the field trial. All selected schools participated in their different programme activities. The student and teacher programmes were both carefully adapted to suit the characteristics of the group as well as the participants' mood and willingness. Consents from parents and all participants were collected and the project was approved by the Swedish Ethical Review Authority, as required by Swedish legislation.

Obstacles: Twelve Schools were selected, and all participated with only one change. One Teacher School and one Control School switched roles in the programme after the final sampling. In one control school, the students did not respond to the second and third questionnaire. In addition, another control school declined to participate in the interviews. Questions to students about student–student relationship and social awareness were not used in Sweden as they were not included in the ethical vetting.

10.4.5. Quality assurance of the project's final deliverables

In summary, the project has managed to complete the deliverables on time with the exception of the field trial in Germany since the German team did not receive permission from the relevant ministry to conduct the field trial in German schools.

The main deliverables are

1. HAND in HAND catalogues: Catalogues for SEI assessment, SEI school staff programmes, and SEI student programmes have been developed and published on the project website.
2. HAND in HAND assessment: Assessment tools to measure Social and emotional competencies and intercultural competencies/ diversity awareness have been developed, both quantitative measures and qualitative measures (semi-structured interviews, focus groups).
3. The HAND in HAND programme for school staff: A programme with a set of learning activities (a combination of personal development activities and classroom-based activities) to increase the Social and emotional competencies and intercultural competencies/ diversity awareness of school staff, including their relational competence, has been developed (Jensen, Gøtzsche, Andersen Réol, Dyrborg Laursen, Lund Nielsen, Denk, Kozina, Vršnik Perše, Marušić, Jugović, Rasmusson & Oskarsson, 2018).
4. The HAND in HAND programme for students: A programme with a set of learning activities to help develop students' Social and emotional competencies and intercultural competencies/ diversity awareness (with a focus on the competencies needed to build an inclusive society) has been developed (Marušić, Puzić, Jugović, Košutić, Matić, Mornar, Lund Nielsen, Jensen, Rasmusson, Oskarsson, Denk, Kozina & Manja Veldin 2018).
5. HAND in HAND guidelines for policy and practice: The results of the field trials, and the policy questionnaire (mapping of national policy contexts regarding Social and emotional competencies and intercultural competencies/ diversity awareness) were published at the end of 2019.

10.5. Discussion

The discussion is organised in line with the three main areas of the quality assurance: the project's planning and goal setting, organisation and execution, and the final deliverables.

10.5.1. Quality assurance of the project's planning and goal setting

The well-structured application and the checklist allowed the participants to obtain an overview of all work packages and all deliverables. Each partner had distinct roles that it made it clear who was responsible for each deliverable. The risk strategy identified several possible risks in the project and needs for things like replacement schools and professional translators were identified. However, it is hard for project members to foresee all possible risks in the planning stage of a project (Williams, 1995). Advice for future projects would be to involve external experts in this process to try to identify and perceive risks. One risk that was not foreseen was that the relevant ministry in Germany denied permission to access the schools and, thus, the planned field trial in Germany was cancelled. Nevertheless, the programmes were implemented and evaluated on a smaller scale in Danish schools instead, during the spring of 2019. Another unforeseen risk was that the tight schedule made it impossible to obtain approval for all of the scales used in the assessment of the students and, therefore, two scales could not be used in Sweden. The other risks listed in the risk management strategy were either not realised or handled by the management strategies. An example is the risk of school dropouts which was managed by the fact that a sampling of replacement schools was made in each country.

10.5.2. Quality assurance of the project's organisation and execution

The HAND in HAND programme is well managed and almost everything has been performed according to the plan outlined in the application. Monthly reports, newsletters and online meetings have ensured that all partners are updated on the stage of the project. The project meetings have been productive and successful and, according to the questionnaires, the participants have reported being satisfied with the meetings.

The partners in Denmark worked with the programme for school staff during the first year of the project in the collaboration with the rest of the partners (see Jensen and Gøtzsche, in Kozina et al., 2019). The student programme was developed by the Croatian partners and in the collaboration with project partners in the same period (see Jugović, Puzić and Mornar, see Kozina et al., 2019). Thereafter, the field trials were carried out during autumn in 2018 at 12 schools in Slovenia, Croatia and Sweden. The risk of not having a stable group of trainers through the

project was solved by ensuring that all school systems had a stable group of teacher trainers who also participated in all of the training arranged by the responsible project partner. Yet, this was not the case for the trainers in the student programme, which might have been preferable. In that way, we could have avoided the trainers being unsure about how the exercises would work out in the student groups. However, this was managed by having cognitive labs in all school systems. By having these labs, the trainers had an opportunity to test some of the exercises and obtain feedback from the students on how they worked out. After the cognitive labs, the results were discussed and adaptations to the exercises were made where necessary. Still, too many different adaptations in the school systems might cause bias in the quasi-experimental design, although keeping the focus on implementation of the core issues reduced this risk. Training for those responsible for the student programme was not planned, even though some training was carried out in the national contexts. This may have had an effect on the delivery of the student programme, for example when it comes to the balance between fidelity and adaptation in the student programme.

10.5.3. Quality assurance of the project's final deliverables

The project has been successful in producing the main deliverables: the three catalogues (SEI assessment, SEI school staff programmes, and SEI student programmes), assessment tools to measure Social and emotional competencies and intercultural competencies/ diversity awareness, the programme for school staff, the programme for students, and the guidelines for policy and practice. In addition, external quality assurance will be provided in the process of publishing results in scientific journals according to the peer review process.

10.6. Conclusion

HAND in HAND is a well-designed and well-managed project. The project has overall met the standard that was initially established. The dropping out of Germany could have been avoided with even more preparation and a longer time frame.

The coordination could have been better between the training of those who were leading the teacher programme and those who were leading the student programme in each country. Greater effort was put into training the persons who delivered the teacher programme in each country than the persons delivering the student programme. Moreover, the mix of fidelity and adaptation was not discussed in the student programme in the same way as in the teacher programme. Like in all projects, the timeframe introduces limits and, if the scales used in the assessment tools had been developed earlier, they could all have possibly been approved by the ethical committee in Sweden.

As mentioned, quality assurance in the HAND in HAND project includes several levels, both the project management level and the implementation of the programmes at the schools. The evaluation of the HAND in HAND field trials is discussed further and the results are outlined by Rožman, Roczen and Vieluf (2020) as well as Vieluf, Denk, Rožman and Roczen (2020).

10.7. References

- Durlak, J. A. (2016) Programme implementation in social and emotional learning: basic issues and research findings, *Cambridge Journal of Education*, 46(3), 333–345.
- Durlak, J. A. (2015) Studying Programme Implementation Is Not Easy but It Is Essential. *Society for Prevention Research*, 16, 1123. Retrieved from: <https://doi-org.proxybib.miun.se/10.1007/s11121-015-0606-3>
- Jensen, H., Gøtzsche, K., Andersen Réol, L., Dyrborg Laursen, H., Lund Nielsen, B., Denk, A., Kozina, A., Vršnik Perše, T., Marušić, I., Jugović, I., Rasmusson, M., & Oskarsson, M. (2018). *HAND in HAND programme for school staff, manual*. Retrieved from www.handinhand.si
- Kozina, A. (Ed) (2020). *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments*

across Europe: Relationships matter. Hamburg, Germany: Verlag Dr. Kovač.

Lund Nielsen, B., Dyrborg Laursen, H., Andersen Reol, L., Jensen, H., Kozina, A., Vidmar, M., Rasmusson, M., Marušić, I., Denk, A., Roczen, N., Jurko, S., & Ojsteršek, A. (2019). Social, emotional and intercultural competencies: a literature review with a particular focus on the school staff. *European Journal of Teacher Education*, 42(3), 41–428.

Marušić, I., Puzić, S., Jugović, I., Košutić, I., Matić, J., Mornar, M., Lund Nielsen, B., Jensen, H., Rasmusson, M., Oskarsson, M., Denk, A., Kozina, A. & Veldin, M. (2018). *HAND in HAND programme for students, manual*. Retrieved from www.handinhand.si.

Oliveira Reis, F. (2009) *Fundamentals of a 'common quality assurance framework' (cqaf) for VET in Europe*. CEDEFOP. Retrieved from: <https://www.cedefop.europa.eu/en/publications-and-resources/publications/5168>

Olsson, A., & Skjöldebrand, C. (2008). Risk management and quality assurance through the food supply chain—case studies in the Swedish food industry. *The Open Food Science Journal*, 2(1).

Research Information Network. (2010). *Quality Assurance and Assessment of Scholarly Research: A Guide for Researchers, Academic Administrators and Librarians*. Research Information Network. Retrieved from: www.rin.ac.uk/quality-assurance

Rožman, M., Roczen, N., & Vieluf, S. (2020). Evaluation of the HAND in HAND programme: Results from questionnaire scales. In A. Kozina (Ed.), *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe* (157-194). Hamburg: Verlag Dr. Kovač.

Vieluf, S. Denk, A. Rozman, M. & Roczen, N. (2020). How do the participants evaluate the HAND in HAND programme? Results of semi-structured focus group interviews. In A. Kozina (eds.). *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe - Relationships matter*, 195-218. Verlag Dr. Kovač.

Williams, T. (1995). A classified bibliography of recent research relating to project risk management. *European journal of operational research*, 85(1), 18–38.

Chapter 11:
Evaluation regarding implementation of the HAND in HAND
programme

Birgitte Lund Nielsen

11.1. Introduction

Strong empirical evidence shows that the way a programme is put into practice, its implementation, is a determinant of its outcomes (Durlak, 2016; Durlak & Du Pre, 2008). Therefore, the evaluation of the HAND in HAND project has also addressed the implementation. The present part of the evaluation report focuses on the methods and results, but does not include the theoretical background concerning implementation issues, which is thoroughly discussed in Nielsen (2020).

11.2. Research aim

The aim of the empirical implementation research in the context of the HAND in HAND programme was to follow over time the implementation in three school systems: Sweden, Slovenia and Croatia.

The research questions guiding the evaluation of the implementation:

- How do the trainers perceive the process of translating the programme to the local conditions?
 - What do they perceive as helpful?
 - Which challenges do they report?
 - What do they report having learned in the process?

The trainers' reflections on translating the programme to local conditions include references to the 'active ingredients' in the HAND in HAND project. These are described in the project materials under the headings of: 1) working with a variety of inner meditative exercises, more outgoing physical exercises and dialogue exercises; 2) the use of 'gearshifts', e.g. between outgoing and more inward going exercises; and by 3) working to establish close contact with school staff and students (Jensen & Goetzsche, 2020).

11.3. Methods

Input from the trainers in the three school systems was systematically collected after each 'session' at the schools, from introductory meetings, to capacity-building in the HAND in HAND programme modules for teachers, school-leaders/counsellors, and students, respectively. An electronic survey instrument was used to frame these structured reflection logs.

11.3.1. Reflection log

The reflection logs included both open-ended questions, e.g. asking for the trainers' experience of the social climate and atmosphere during the session, and Likert-scale questions, e.g. about the experience of success with respect to the 'active ingredients' mentioned above.

All the questions follow here:

Please fill in this questionnaire immediately after each Hand in Hand session (teaching day); both introduction meeting, teacher/student training etc. Use English when answering in the open categories.

1. Name of trainer

2. Country (closed, single choice: Slovenia, Croatia, Sweden)

3. Type of session (closed, single choice, category 4 with comments: Student training, Teacher training, Leader/counsellor training, Other, write what)

4. Date, and number of training (in the row of sessions, e.g. "day 1 for teachers") (open)

5. Describe shortly, in your own words, the agenda and the overall aims of the session. Include if any planned adaptations have been made in relation to the manual (open)

6. What went particularly well in the session? (open)

7. What was not well received? / What was rather difficult? (open)

8. Questions about timing, the plan and organization (5-point Likert, from 'to a very high degree' to 'to a very low degree', plus 'don't know/non-eligible')

- Regarding the timing, to what degree was it possible to run the training session according to schedule?*
- Regarding the taught content, to what degree was it possible to run the training session according to plan?*
- Regarding organizational aspects (logistics, materials, the physical environment etc.) to what degree was it possible to run the lesson according to plan?*

Notes/comments on timing etc. (open)

9. Core components/active ingredients (5-point Likert, from 'to a very high degree' to 'to a very low degree', plus don't know/non-eligible')

- *To what degree, did you experience to succeed in including a variation with different kind of exercises (dialogue, physical exercises, inner exercises)?*
- *To what degree, did you experience to succeed in using "gearshifts" (e.g. between outer going and more inward going exercises)?*
- *To what degree, did you experience to succeed in being in close contact/dialogue with the participants along the session?*

Notes/comments on core components/active ingredients (open)

10. Describe your experience of the social climate/atmosphere during the session. Use examples if you can. (open)

11. Overall, how was the session received according to your impression? (5-point Likert, from 'very well' to 'not well at all')

Elaborate on your answer (open)

12. What are the most important learning/insights you as a trainer take with you from this session? (open)

13. What adaptations in the programme would you suggest looking forward? (open)

11.4. Analyses

The inputs in the electronic reflection log (n=121) covering the period from May to December 2018 were analysed by inductive thematic analysis (Braun & Clarke, 2006) of open reflections, and frequency analysis and cross tabulations of the Likert-scale answers.

11.5. Findings

A glance at the rich data set showed that, while it was clear that some overarching issues could be identified, differences also existed from school to school, from country to country and not least depending on whether the trainers were working with school students, teachers or school-leaders/counsellors. Some of the findings below are therefore organised according to the latter structure, but are also discussed with respect to some of the dominant themes.

11.5.1. Trainers' experiences related to the active ingredients

First, we present some findings from the Likert-scale questions about the trainers' experiences of succeeding with the active ingredients identified above when running the programme in the local conditions. Results are given for all answers summed across the session types and school systems with the answers after the first 3 months in the phase with sessions at schools (May, June and July 2019) shown in Figure 11.1., and the answers from the full dataset in Figure 11.2.

We need to be careful with the interpretation. The two representations are not directly comparable, e.g. when it comes to a division into school systems and persons in the inputs for the reflection log, but it is any case interesting to highlight some tendencies. While looking at the data, over time there appears to have been a development whereby the trainers started to feel more confident about having succeeded when leaving a specific session. In Figure 11.1., one sees 26%–35% answering to a high or very high degree, but 51%–65% in Figure 11.2. It is also important to emphasise that it is the particular module the trainers just finished which they are reflecting on in the log, not the full programme.

This demonstrates how the trainers grow to become more confident and it might also be cautiously inferred that they, and perhaps also the school staff and students, developed a deeper understanding of the programme's core ingredients. Questions about achieving a more developed understanding and more confidence over time are discussed below in the section on qualitative data.

Figure 11.1. also shows that in the initial sessions the challenges were particularly related to the lack of feeling of being in close contact (the first question in Figure 11.2.). The inputs for the log in the first months show that 31% experienced having succeeded in establishing close contact with the participants to a low or very low degree. In the full dataset (Figure 11.2.), the answers across the three questions are more alike²⁶.

²⁶ For information, the non-eligible answers in Figure 1 in particular are about not having worked very much on these elements in the first sessions.

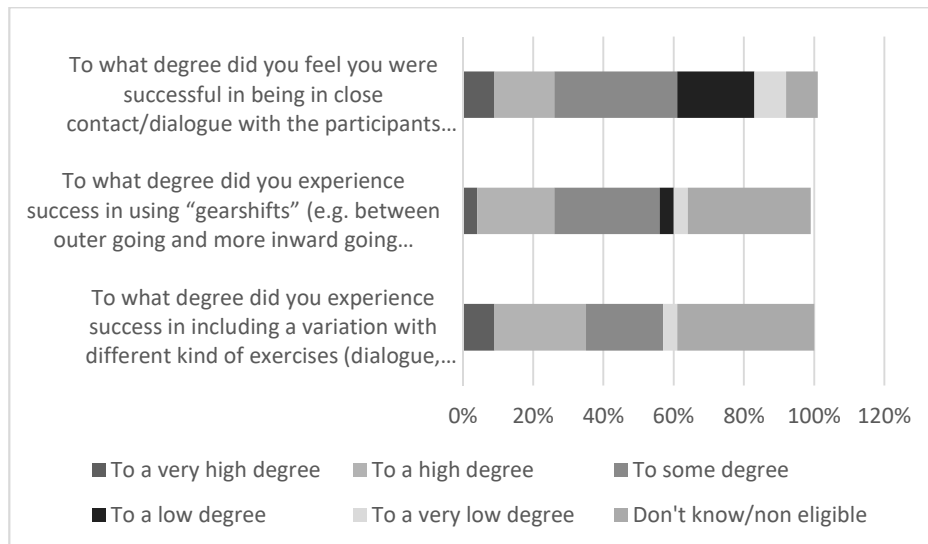


Figure 11.1. Likert-scale answers about the 'active ingredients' after 3 months (July, 2018)

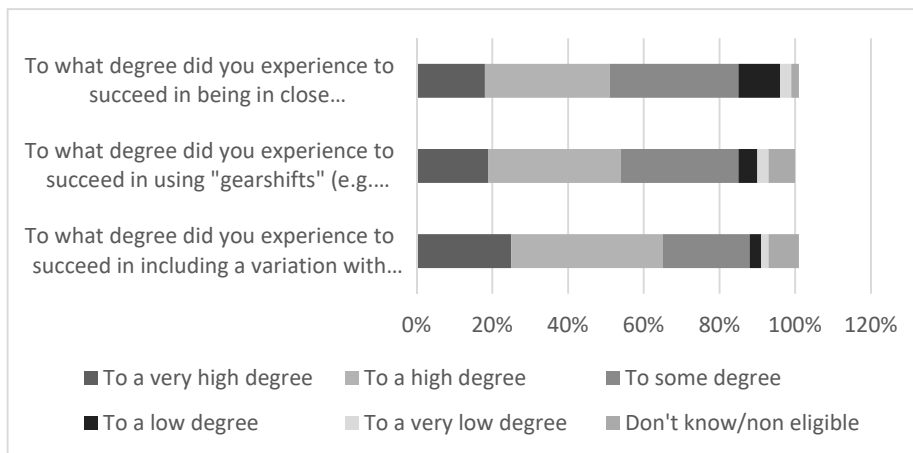


Figure 11.2. Likert-scale answers about the 'active ingredients' with all answers (n=119)

Crossing the full dataset with session type and participant group confirmed that the challenges were especially experienced in the initial sessions held at the schools. Interestingly, the 51% of respondents who answered “to a high” or “very high” degree to the question about close contact (Figure 11.2.) covers some of the differences depending on participant group, i.e. 59% in relation to student training sessions vs. 41% to the teacher training sessions. Here, it is again important to be cautious in making conclusions since different trainers were typically involved with the various groups. There are some national issues, but we are also cautious while interpreting these differences since there are also differences from school to school in each country. This shows the situated and strongly context-specific nature of these social meetings during the HAND in HAND modules.

In summary, there was overall development in time with the trainers' experiences of succeeding with respect to all three areas of the 'active ingredients', including apparently 'solving' some of the challenges related to the lack of close contact felt in the first sessions.

11.5.2. Interplay between the trainers and the school staff

The development in relation to the experience of being in close contact can be further illustrated with the open reflections of the trainers, and how they developed over time. Table 11.1. gives an example of reflections from the same trainer over time, referring to sessions with teachers.

Table 11.1.

Reflections of one trainer from one country referring to sessions with teachers. Example of development over 7 months.

Date	Quotes from reflection log
June, 2018	...the atmosphere was bad; some teachers did not say hello back ...it felt like they were forced to be there by school coordinator. Their thoughts were somewhere else and it was hard to be enthusiastic about the programme and the project....
August 2018	The atmosphere was better, a lot of positive feedback ... participants started to share their experiences, thoughts. Some reflected that they were really focused on finding solutions on the first day and they feel now ... they are here just for themselves. They provided some insight at the end on how and what they find useful and were eager to use some also with students.
September, 2018	The atmosphere was very positive even though the teachers came after their classes and were tired they were in a good mood ...especially after the round, the connection was felt. By the fact that they had used several activities on their own, it felt that the programme was positively received. And that we are a group now.
December 2018	The climate was positive, accepting, it is also a result of the last module and we know each other well by now. It felt that the participants are relaxed to share their opinions, thoughts. Also, at the end, there was a lot of gratitude and hugs and connection felt in the room.

Development over time is illustrated by the open reflections (Table 11.1.). This is seen in the way the development in the group of teachers is presented through the eyes of the trainer, realising a lack of ownership of the programme among the

teachers from the outset, and later acknowledging the teachers' growing openness, and that they were contributing with their own experiences. Yet, there is also an implicit development over time, from the trainer being descriptive: "*the teachers did not say hello*", towards highlighting more dynamic issues in the interpersonal relations, e.g. the teachers' experiments in their own classes and their contributions to the co-creation: "*we are a group now*".

Here is an example taken from one of the other school systems, also with regard to teacher training:

The teachers were worried about the days in December when they usually have a lot of work. There are many challenges for teachers with the students and I am not sure how we can handle this in the teacher training. (June 2018)

It was much easier to be in quite close contact with the teachers this second time, and added in relation to adaptation: We did not always follow our schedule, but instead shifted exercises when we felt that the teachers needed that. (October 2018)

There was a special, shorter 2-day programme for the school-leaders and counsellors. Based on the reflections in the log, particular issues were at stake with this participant group. The programme's value is, for instance, discussed more at a meta school-development level, as shown in these two reflections (from two different school systems):

The atmosphere was pleasant, but a bit reserved, as if the participants were not fully convinced of the value of this kind of programme. They could not see the relationship ...with the quality...teachers deliver in the classroom.

The climate was changing during the day. We had a positive atmosphere for most of the time and the participants opened up and talked about their experiences. At some point, it was a bit of a struggle when the two principals questioned the theory and many of the others did not agree. It was good to have that discussion.

The complexity entailed in understanding the sessions with this group also concerns the group's heterogeneity. The mentioned meta school-development perspective might be a typical school-leader perspective. In general, the agenda of school leadership is not always the same as the agenda of teachers when talking of professional learning activities, as also mirrored in this dataset, e.g. in the reflection about who made the decision for the school's participation (Table 11.1.).

Moreover, there are differences from country to country in who is participating in this group. In one country, health counsellors for example were included in this short programme. They can have quite a different agenda than teachers and school, as illustrated here:

The staff from the student health teams were happy that there, for once, was focus on the students' well-being.

The quote refers to an experience of a health counsellor that an agenda they apparently tried to raise is now being raised by the broader group of school staff.

11.5.3. Interplay between the trainers and the students

The reflections of the trainers working with modules for school students reveal some of the same issues like with the teachers, along with other kinds of issues. Many of the reflections about what went well in the modules concern students being active, interested, engaged, curious etc., not far from the reflections acknowledging teachers' active contributions. However, certain issues also arise in relation to, e.g. classroom management:

Students participated and were engaged in all activities, however, as a group are quite loud and sometimes difficult to maintain their focus....

...there were a few students that were disturbing most of the exercises.

Yet, the reflections made by the trainers over time show a willingness to take a student perspective by realising the complexity of everyday life at school as experienced by a student, and that this can affect the students' engagement with the programme activities:

Today, the students seemed to be under stress, probably due to tests and grading. They seemed a bit uninspired and, when some girls refused to do the exercise with the chair, others followed.

Hence, while it is mirrored in the trainers' reflections that the students' mood changed from session to session, as one trainer put it, one can also identify some kind of development over time. Here are some quotes for illustration:

The session as a whole went well, much better than module 2. Already when we entered the school building some students were there and were excited to see us and were looking forward to what we would be doing ... students came and eagerly volunteered to help (we had two boxes of yoga mats with us).

This is a large student group with a wide variety of different students and attitudes. Nevertheless, they are successful in listening to each other and co-operating. Some students who did not want to participate earlier showed some curiosity today and partly participated in the exercises. It seems that they are starting to realise that it's voluntary and that it's perfectly ok to attend according to one's own ability.

11.5.4. Reflections on own learning during the implementation at schools

One of the last questions in the reflection log concerns the most important learning insights the trainers themselves had experienced from the session. Some of these reflections are connected to specific issues raised elsewhere in the same input for the log, like the reflections revealing frustration at the beginning of the programme described above, but also the realisation when better contact was established:

On one hand, I wonder what I bring to these sessions that are so difficult, am I not as prepared, engaged ...I can only expect very small steps.

I don't know what made the difference in them being able to participate better in this first hour of the session – was it something we did or it was just coincidence?.

...it takes some time to establish good contact with different students and groups of students. Now we feel that they are more relaxed and that they dare to trust us....

Some overarching themes identified in the thematic analysis of the trainers' reflections in this part of the logs are listed and exemplified in Table 11.2. Note that there are both themes independent of the participant group and other themes related to the respective groups (Table 11.2.).

Table 11.2.

Reflections from the logs concerning the question: “What are the most important learning/insights you as a trainer take away with you from this session?”.

Theme	Quotes to exemplify the theme
Building trust over time (all participant groups)	<ul style="list-style-type: none"> • ... it takes some time to establish good contact with different students and groups of students. Now we feel that they are more relaxed and that they dare to trust us. This seems to be especially true for the ‘cooler’ guys. • Creation of an atmosphere of mutual trust, support and authenticity is the most important element for the success of this programme. • That being in close contact is not always easy in all groups and it sometimes becomes easier with more time.
Adaptation, as each class and group of teachers is “it’s one” (all participant groups)	<ul style="list-style-type: none"> • You do not have to do everything that is planned. It is better to address one idea in such a way that it gets through. • ...it is important to listen to the group and make adjustments accordingly. We have three different classes and we make small adjustments so as to make it work for the students.
Own agency (all participant groups)	<ul style="list-style-type: none"> • That I can do it. That it was possible to lead a group and have the gearshift in mind. That the exercises are well accepted, even in the leaders’ group. • I can stay calm even in such difficult situations when students are not participating and responding to my questions. I have a strategy. • I got a sense I can really follow the students' energy and (lack of) of focus and respond so that I lead activities in a way that helps them use energy, restore focus or bring awareness inwards (depending on what they need). This brings me a sense of inner satisfaction and gratefulness and humbleness to be able to do this.
Physical environment (all participant groups)	<ul style="list-style-type: none"> • ...it is important to have a room where you are able to move around as well as to sit and talk.
Co-reflection with students to understand and support instead of blame (student programme)	<ul style="list-style-type: none"> • ...at first, I was getting annoyed ... thinking why ... not follow the instructions ... then it hit me ... it is too challenging. Once I had this acceptance and compassion, everything was easy. We reflected together on how this was difficult for them and verbalised strategies that would help...so, my insight was – do not judge, blame, try to understand ... inquire about it. • Make the meaning of the exercises clear to the students.

Co-reflection with school professionals in relation to the student programme	<ul style="list-style-type: none"> • We are making progress in students' ability to reflect.... The school coordinator also commented that we are having an important impact on one particular student who is responding very well to the activities in the module (i.e. she is opening up). • To establish good relations with teachers and other school staff. • The importance of having a teacher or another person who knows the group involved.
Hard to explain (school staff programme)	<ul style="list-style-type: none"> • That it is hard to explain the approach ... the thoughts behind relational competence, the 'new' way of seeing children as social beings.
Professional agency (teacher programme)	<ul style="list-style-type: none"> • The idea of empowering their own capacities was new....
Relations with co-trainers	<ul style="list-style-type: none"> • It feels safe working together and we can take turns and help each other. • The team of trainers worked well. • Turn for help to your partner, when needed.

11.6. Discussion and perspectives

An initial reflection based on the findings is that the high complexity of the HAND in HAND project and of implementation processes generally is confirmed. Accordingly, the answer to how the trainers perceived the process of translating the programme to the local conditions is not simple. The reflection logs contain many indications of challenges, especially in the trainers' first meetings with the participants. However, there seems to be a development over time whereby the trainers generally grow to become more confident, feeling that the collaborative and active work with the ideas in the project can make a difference. The trainers appear to appreciate what the meetings with the participants do to themselves personally/professionally. The data indicate that over time most trainers developed a level of professional agency (Edwards, 2009) in relation to working with Social and emotional competencies and intercultural competencies/ diversity awareness in a concrete school setting. This appears to be a two-way transition process in the interplay of trainers and participants with relationship building (contact and trust) as a central aspect. The professional agency indicated in the data therefore appears to be very much about the relational aspects, the capacity to work with school staff and co-trainers drawing on distributed resources, and translating the programme content in a meeting with participants while also acknowledging their perspectives and contributions. Hence, the perceived learning outcome from the trainers, outcomes that seem to (slightly) change the

trainers' sense of professional self and social self, but often in a process with some struggles, clearly refer to one particular aspect of the HAND in HAND programme: the question about relationship-building. Nevertheless, it may be argued that the findings have a more generic bearing when it comes to school development. Referring to professional agency, and to Social and emotional competencies and intercultural competencies/ diversity awareness, it may be argued that the building of trust and relations illustrated here is a prerequisite for supporting teachers' professional learning also *beyond* a programme where the content is about i.e. relationships.

Table 11.1. highlights the teachers' eagerness to use what they did in the training also with the students through the eyes of the trainer. This leads to another key reflection, including findings from another part of the qualitative data. Vieluf, Denk, Rozman and Roczen (2020) stress that participating teachers appreciated the atmosphere during the training and the opportunities for personal development and self-empowerment, but felt unsure about how to connect the things they had learned with their everyday work. The issue about including even more enactments and collaborative reflection on these during work in the programme was discussed early on in the implementation, in relation to both the programme for trainers and teachers, but it proved to be too hard to arrange due to e.g. practical reasons and experimental conditions. The reflections over time given by the trainers however show the importance of one's own consciously reflected practice in developing professional agency (the trainers), and this is surely also an issue for teachers who are novices in the field, while all the trainers had research-based insight into and experience with some of the SEI perspectives before the project. There is no simple solution to this because, as indicated in Table 11.1 and by Vieluf et al. (2020), the teachers also appreciated that they were "*here just for themselves*". It may be highlighted as a more generic dilemma related to school development that the Dewey perspective of learning by inquiry (enactment and reflection intertwined) is certainly central to professional learning. But we also need to be aware that we are living at a time when new forms of public management are challenging professionals' judgement and autonomy, with top-down demands being the 'new normal'. Hence, the experience of taking a step back and doing something for yourself might be a new positive experience for a teacher.

This dilemma leads to the third key reflection. Although development over time is mirrored in the trainers' reflections, this is certainly not a straightforward process. Instead, the complexity of adaptive processes is illustrated. Røvik (2016) stresses that new ideas typically trigger complex processes involving sense-

making and the elaboration of meaning (over time), but also power plays, resistance and negotiation. This describes quite well the trainers' overall experiences. Other scholars have noted that many innovation projects, like the HAND in HAND programme, are based on a rational planning approach with expert-driven designs being implemented, but emphasising that the assumptions underlying rational planning are inconsistent with complex adaptive systems. Complex systems are inherently non-linear and exhibit a great deal of noise, tension and fluctuation in interaction with the rest of the environment. A provocative question here is whether the whole idea of universal school development programmes adaptable for all contexts is simply an illusion. This would be a misinterpretation. We as researchers and professionals must be able to share and cooperate to develop pedagogy across schools and school systems, i.e. in the crucial field of Social and emotional competencies and intercultural competencies/ diversity awareness. But we must carefully consider how to develop a positive system-level change. Darling-Hammond (2005) illustrates how educational change generally depends on initiatives at different levels of the system, and that most successfully implemented reform initiatives are those that induce top-down support and the input of new ideas at the same time supporting bottom-up development. Downes (2014) highlights the need to examine multiperson systems of interactions when analysing the effects of reform initiatives. The data from the HAND in HAND project confirm the need for such a system-level view, in relation to both the implementation of reform initiatives, as stressed by Darling-Hammond (2005), and in relation to research looking at the implementation, where Downes (2014) discusses how to understand system change and emphasises e.g. the need to examine the two-way flow in a system of reciprocity to incorporate feedback. He also suggests a dynamic system theoretical framework that also highlights individual responsibility within the totality of the system (Downes, 2014). The analyses of the implementation data presented in this chapter provide an insight into the reciprocal interaction between the trainers, their development of individual and relational agency, and the system/subsystems that framed the implementation of the HAND in HAND project.

11.7. Concluding remarks

Summing up, the development over time in the trainers' confidence in relation to working in schools with the HAND in HAND programme is revealed in this chapter. It entails a multifaceted and sometimes quite challenging process of

professional learning and of developing professional agency. The need for a similar process for teachers to develop confidence over time by applying the HAND in HAND approaches in their own classrooms can be hypothesised.

The identified challenges especially concern the trainers establishing close contact and trust in the participant groups. Based on the quantitative and qualitative data from the implementation survey there appears to have been a transition process over time with relationship building between trainers and participants. The trainers stress that it takes time to create an atmosphere of mutual trust. They refer to their own learning insights in relation to helping and supporting each other in the team of trainers to meet the challenges. This indicates the development of relational agency. While working with the school staff particular issues have been about supporting teachers in the feeling of participating ‘for themselves’, not just on a top-down decision from e.g. the leader. The trainers refer to their own learning insights in connection to the idea of empowering teachers’ capacities.

Working with students has for example included issues about classroom management. The trainers refer to their own learning insights with respect to a nuanced understanding of the challenges experienced by the students.

Looking then at the continuing process of spreading the results from the Hand in Hand project it should be considered how the materials developed in the project can be shared in a balanced manner. This means presenting the idea of Social and emotional competencies and intercultural competencies/ diversity awareness and the crucial active ingredients from the HAND in HAND project, and the need for external supervisors to support a whole-school process, while also highlighting the importance of adapting to and acknowledging the local context, competencies and professional agency.

11.8. References

- Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 77-101.
- Darling-Hammond, L. (2005). Policy and change: Getting beyond bureaucracy. In A. Hargreaves (Eds), *Extending educational change*, 362-387. Dordrecht: Springer.
- Downes, P. (2014). Conceptual Framework and Agenda: Beyond Bronfenbrenner (1979, 1995) to Interrogation of Blocked Systems via Structural

- Indicators. In P. Downes (Eds). *Access to Education in Europe: A Framework and Agenda for System Change*, 29-45. Dordrecht: Springer.
- Durlak, J.A. (2016). Programme implementation in social and emotional learning: basic issues and research findings. *Cambridge Journal of Education*, 46(3), 333-345.
- Durlak, J.A. & DuPre, E.P. (2008). Implementation Matters: A Review of Research on the Influence of Implementation on Programme Outcomes and the Factors Affecting Implementation. *American Journal of Community Psychology* 41, 327–350.
- Edwards, A. (2009). Relational agency in collaborations for the well-being of children and young people. *Journal of Children's Services*, 4(1), 33-43.
- Jensen, H. & Goetzsche, K. (2020). Development of the social, emotional and intercultural learning programme for school staff. In A. Kozina (eds.). *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe - Relationships matter*, 83-106. Verlag Dr. Kovač.
- Nielsen, B.L. (2020). Implementing the HAND in HAND programme for school staff and students. In A. Kozina (eds.). *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe - Relationships matter*, 107-130. Verlag Dr. Kovač.
- Røvik, K.A. (2016). Knowledge Transfer as Translation: Review and Elements of an Instrumental Theory. *International Journal of Management Reviews* 18, 290–310.
- Vieluf, S. Denk, A. Rozman, M. & Roczen, N. (2020). How do the participants evaluate the HAND in HAND programme? Results of semi-structured focus group interviews. In A. Kozina (eds.). *Social and emotional competencies and intercultural competencies/diversity awarenessfor inclusive school environments across Europe - Relationships matter*, 195-218. Verlag Dr. Kovač.

PART D: SUMMARY AND INTEGRATION OF FINDINGS

Chapter 12: How effective are the HAND in HAND Programmes and how could they be improved for future use?

Svenja Vieluf, Nina Roczen, Mojca Rožman

12.1. Introduction

The main aim of the HAND in HAND programmes was supporting the development of more inclusive classrooms, in which every student feels accepted and able to achieve their potential, by fostering the social, emotional and intercultural (SEI) competencies of students and school staff. The external summative outcome evaluation used a randomized control-group experiment to assess with multiple methods to what extent the programmes had succeeded in achieving these aims during a field trial with 36 schools/classrooms (one classroom per school) set in three different education systems (the Croatian, the Slovenian, and the Swedish system). Additionally, the external summative outcome evaluation aimed at understanding the perspectives of participants on the quality of the programme. To this end, participants were asked during semi-structured focus-group interviews what they liked and disliked about the programmes and what they think they learned from participation in the programmes. Hence, a strict experimental study that allowed for detecting causal effects of the programme was combined with a strategy that recognized the importance of the perspectives of participants and that also allowed for identifying possible unintended effects. The external summative outcome evaluation was further complemented with an external formative outcome evaluation, which aimed at identifying possibilities to improve the project outcome, i.e., the HAND in HAND programmes for students and school staff. Additionally, the project was also subject to two different internal evaluations: The first was an internal summative process evaluation that aimed at understanding how the programmes had been implemented by the trainers, what had worked well and what challenges they had encountered. The second was an internal formative process evaluation that had the aim to continuously monitor the overall project management and the implementation of the programmes in schools during the project and to give frequent feedback in order to foresee difficulties and help making adjustments early in time.

Hence, different evaluation components were internal and external, had a summative and formative purpose and concerned outcomes and processes. The present chapter has the aim to integrate findings and insights stemming from these multifaceted evaluation approaches in order to draw conclusions about the overall quality of the HAND in HAND programmes and their implementation during the field trial as well as possible directions for future development of the programmes.

12.2. How effective are the HAND in HAND programmes? Main results from the summative evaluation

The external summative outcome evaluation of the HAND in HAND programmes examined whether the programmes served to improve the social and emotional competencies and intercultural competencies/diversity awareness of students and school staff as well as the classroom climates and school climates in participating classes/schools during a field-trial in 36 classes in 3 different school systems (12 classes in Croatia, 12 classes in Slovenia, and 12 classes in Sweden). To this end, changes between three measurement points, more precisely between T1 and T2 for short-term effects and between T1 and T3 for middle-term effects (T1 - before the training, T2 - after the training and T3 - follow-up) in several sub-dimensions of the CASEL components (self-awareness, self-management, social awareness, relationship skills, responsible decision-making; CASEL, 2003) as well as in intercultural competence/diversity awareness were examined using self-report- and vignette-based measures. Changes in the quality of the classroom climates between the three measurement points were assessed through repeated use of questionnaire scales concerning the perceived quality of classroom climates, through repeated network analysis, and through a qualitative content-analysis of respondents' responses to interview questions about changes in classroom and school climates. Finally, participants were also asked directly during semi-structured focus group interviews, what they thought they learned through participation in the HAND in HAND programmes and how they evaluated the programmes themselves.

Evidence concerning causal effects of the HAND in HAND programmes on social and emotional competencies and intercultural competencies/diversity awareness is mixed. Some of the expected differences between experimental groups with regard to self-report scales were statistically significant in some of the school systems, but the effect sizes were small, and we observed almost as many significant effects that had a direction opposite from what we had expected (i.e. a change for the worse in the classroom climate of a class that had participated in the programme). Above all, the effects (both, the expected and the unexpected effects) were often inconsistent across experimental groups and school types.

A special feature of the HAND in HAND evaluation design was that it compared three different experimental groups with a control group: one experimental group where only a student programme had been implemented (group B), one

experimental group where only school staff programmes had been implemented (group C), and one experimental group where all programmes had been implemented (group D). The aim of this design was to test whether a “whole-school approach” was more effective than programmes addressing only one type of stakeholder in each respective school. Hence, we expected to find effects on students either in groups B and D or only in group D (plus additionally an indirect effect in group C – in case that the teachers had implemented the exercises learned during their programme with their students). We further expected effects on school staff either in groups C and D or only in group D (an indirect effect in group B was, however, also considered possible; this might have occurred when a change in student behaviours caused by the student programme also affected their teachers through interactive processes in the classroom). Yet, we found only few effects that were consistent over groups B and D or over groups C and D (or over all three conditions) and none of these consistent effects showed up in more than one of the three school systems. As compared to the consistent effects, we found more significant effects for group D alone, but also none of these effects was consistent across school systems. Support for a whole-school approach was observed only in Slovenia (a number of differences between group D and the control group are significant here), but not in the other two school systems.

The result-pattern most in accordance with prior expectations about programme-effects was observed for the scale *observe*. This scale assessed the observing, noticing or attending to various stimuli, including internal phenomena (cognitions, bodily sensations) and external phenomena (sounds, smells), which is part of mindfulness and, thus, a central focus of the HAND in HAND programme. 12 out of the 30 possible effects on this scale were significant and positive, implying that there was a stronger increase or a weaker decrease in the ability to observe in one of the experimental groups as compared to the control group. Yet, even for this outcome, the effects were neither consistent across different experimental groups that had participated in a similar programme nor across school systems.

Consistent with the results based on self-report measures, we observed hardly any changes in participants’ interpretations of a social situation, which triggered beliefs about the social difference category “migration background”, or in their repertoire of strategies for dealing with this situation. Hence, content analysis of vignette responses for two time points (T1 and T3) did not provide evidence that

the Hand in Hand programmes increased (neither decreased) participants' intercultural competence/diversity awareness.

In conclusion, our findings from a multiple-method assessment of changes in social and emotional competencies as well as intercultural competence/diversity awareness in different experimental groups do not support the existence of a consistent causal effect of the programme on these competencies. There is some indication that the programme might improve the ability of observing non-judgementally across experimental groups and school types, but more research is needed to corroborate this impression, because not even this effect was consistent across relevant experimental groups and school systems. There is also no evidence of a negative effect of the HAND in HAND programme.

Participants' responses to the interview-questions what they learned during the programme and what they would highlight as perceived outcomes from participating in the programme activities contradict the findings from the experimental study. Many teachers, school leaders and other school staff evaluated the programmes positively and also a number of students liked the programmes. Participants found the programmes "fun", said that it made them feel good and that it created a positive atmosphere in the group. Teachers further liked that they learned concrete techniques for their individual stress-management as well as exercises they can implement in the classroom. Many found this useful. Moreover, many students, teachers and other school staff who had participated in a HAND in HAND programme reported during the semi-structured focus-group interviews that they had acquired social, emotional and/or inter/transcultural competencies in the programme. Self-management competencies were mentioned particularly often and relationship skills were also relatively frequent.

Several reasons for the inconsistency between the randomized-control group study vs. the results of semi-structured focus-group interviews are possible: First, participants (and in particular the adult participants) might have been hesitant to tell the project teams that they think they learned nothing during the programme. Hence, the interview results might have been distorted by social desirability effects. Second, and on the contrary, it is also possible that the questionnaire scales and vignette approach were both not "instruction sensitive". Both instruments might not have been appropriate for capturing the nuances in thoughts, feelings and behaviours that a programme like the HAND in HAND programme causes in a short time period of just a few months (For a more detailed explanation of the concept of *instruction sensitivity* see Naumann, Hartig, & Hochweber, 2017; Naumann, Hochweber, & Klieme, 2016; Naumann, Rieser,

Musow, Hochweber, & Hartig, 2019). What speaks against this latter interpretation is that we observed no effects using different methods (self-reports, other reports and vignettes). Also several small changes were actually observed in participants' responses between the measurement points these were, however, mostly not in accordance with our hypotheses. A third possible explanation is that the questionnaires and vignettes measured outcomes that were slightly different in type from those referred to during the interviews. It is noticeable that many participants described during the semi-structured focus-group interviews that they learned norms regarding social behaviour and relations to self (e.g., "*we should not have prejudices, we should not be racist*") or that they had acquired knowledge about social or emotional processes (e.g., "*how oxygenation is related to learning*"), whereas they mentioned rather seldom that the way they thought about social events or the way they actually behaved had changed. The questionnaires and even the vignette-based measure, in contrast, assessed changes to attitudes and/or behaviours. For example, during the interviews teachers said that they learned "*how body and psyche are linked*" and one questionnaire item asked: "*I pay attention to whether my muscles are tense or relaxed*". Knowing in theory about the connection between body and psyche does not necessarily imply that one becomes aware of the own body reactions in situations of distress. Hence, reporting the first during the semi-structured focus-group interview, but answering with "no" to the questionnaire question is not necessarily contradictory. Questionnaire items and interview responses might concern different stages of change.

Integrating the findings based on self-reports, vignettes and semi-structured focus-group interviews suggests that a majority of participants liked the programmes and found them useful and that they might have learned knew norms and some theoretical knowledge about social and emotional processes and about diversity, but that they did not change their beliefs, attitudes or behaviours. Research on the effectiveness of trainings (more precisely, trainings in the field of teacher professional development, e.g. Guskey, 2000) has shown that the first and most basic stage of change is reached when participants' are satisfied with and experience relevance of the training. This is the case for the HAND in HAND programmes: many participants (a majority of interviewed teachers, school leaders and other school staff, but also many of the interviewed students) said during the semi-structured focus-group interviews that they liked the programme and found it useful. The first stage provides the basis and motivation for building knowledge and changing convictions in a second stage. Results from the

interviews suggest that the HAND in HAND programme has also triggered some second-stage changes. However, only at a further stage can changes in behaviour take place.

A lack of change in student and teacher behaviours also explains why we found hardly any significant effects on classroom climate measures. In the theoretical model presented in Chapter 1, the changes in classroom climates are the result of changes in teachers' and students' behaviours. If these behaviours did not change, then we can also not observe effects on classroom climates. It should, however, be noted that some short term changes to group climates were observed: about one-third of the teacher-interview-groups²⁷ explicitly noted the positive atmosphere during the trainings and also about a third of teacher groups as well as about one-third of the student groups and about 12% of the school leader/other school staff groups who participated in a HAND in HAND programme said that at the end of training sessions participants felt good – even after a long day. Moreover, there is some indication of a positive effect of the training on the relation between adult participants in the HAND in HAND staff trainings: During the semi-structured focus-group interviews, teachers who had participated in the teacher programme reported that they had bonded as a result of their common experience, that they had become closer and cooperated better after the training. Hence, the HAND in HAND programmes appear to have succeeded in creating a momentary positive group-atmosphere during the training and in helping teacher build stronger connections amongst each other, even though it has not triggered more long-term changes to the classroom climates.

12.2.1. Strengths and Limitations of the Summative Evaluation

The HAND in HAND evaluation had a complex design. It combined a randomized control-group experiment with qualitative content analysis of semi-structured interviews. The experimental approach allows for drawing causal conclusions. The interview-approach allows for identifying additional unintended effects and gives those whom the programme is intended to serve a voice. Changes in outcomes of the HAND in HAND programmes were assessed with self-report questionnaire scales, as in many other evaluations, but also with other-reports, vignettes, sociometry and semi-structured focus-group interviews. Using multiple methods for assessing the same outcomes increases the validity of

²⁷ Please note that the groups that were interviewed only represent a sub-sample of the participants.

conclusions (Campbell & Fiske, 1959). Hence, the evaluation helped gaining a comprehensive understanding of programme effects – or a lack thereof. Yet, the evaluation also has some limitations: First, the programmes targeted individual competencies, but also a change in school cultures. Allocation to experimental groups was at the school-level. At the school level, however, the sample size was very small. Therefore, third variables at this level might have confounded the results. The small sample size at the school level also meant that we either had to carry out analysis only at the individual level or at the school level but for all school systems together – in spite of evidence for effect-heterogeneity at this level. Moreover, also the size of the teacher- and school-leader/other school staff-samples were rather small, so that we had to combine both groups for analysis even though their programmes differed considerably in length. A further issue is the high drop-out rate of schools in Sweden.

12.2.2. Conclusions Regarding the Programme's Effectiveness

The main question of the summative evaluation is, how effective the HAND in HAND programme is. The results suggest that the programme has triggered stage-one and possibly also some stage-two changes – i.e. a positive stance towards the programme and motivation to participate as well as changes to attitudes – but only few stage-three changes, i.e., changes in behaviours. Participants liked the programmes and found them useful. They also said they learned something from it. In particular, they have developed an idea about how to approach difficult social situations. However, this does not seem to be sufficient to trigger stage-3 changes of their everyday behaviour. Accordingly, also the classroom climates have not changed as a consequence of the programme. There are a few exceptions: The programme appears to have helped a number of individuals in different age groups and different school systems to become more mindful in terms of observing without judging. Also, the programmes succeeded in creating a positive group-atmosphere during the training itself. Moreover, there is evidence that the programmes have helped teachers to develop a closer connection.

12.3. How can the programmes be improved? Main results from the formative evaluations

One main suggestion for improvement made by the participants themselves is extending the programme. The programme for school leaders and other school staff was only 2 days long, the programme for teachers was 5 days long and that for students encompassed five 90-minute sessions. Many participants found this too short. The shortness could also be a reason for the lack of causal effects of the programmes on intended outcomes. This argument is also supported by the reflection logs for trainers: The trainers described several difficulties during the first sessions and how it took time to build up relationships. Once, the relationships had been established and trainers and participants had attuned to each other, there were only few sessions or even none left for more intense learning. One reason for the shortness of the programmes was the difficulty to get more funding. One hope was that teachers would use exercises they had learned during the programme in their classrooms so that they would practice regularly even after the programmes had ended. However, the interviews suggest that only few teachers did so on a regular basis. Further, many teachers asked for more support for implementation. So the teacher programme appears to have not provided enough support for enabling teachers to consolidate practices of social, emotional and transcultural learning in their classrooms. Hence, for increasing effectiveness and sustainability of the HAND in HAND programmes it might be helpful to extend the programmes and, in particular, to add elements that support teachers with the application and implementation of the programme in the classroom.

How could the programmes better support regular practice and implementation? Participants themselves suggested that it would help if trainers provided guidelines for implementation. Others suggested supervision. Additionally, peer supervision and peer coaching including mutual classroom observations could be helpful (see e.g., Bowman & McCormick, 2000; Glatthorn, 1987; Hargreaves & Dawe, 1990; Showers 1984). Such support for implementation should also aim at supporting teachers' feelings of agency, as the reflection logs for trainers indicate. With regard to support for practicing inner exercises on a regular basis, teachers asked for regular reminders. To this end, email reminders or a mobile phone app could be offered. However, support for regular practice should not result in pressure: It should also be considered that not everybody feels comfortable with

a mindfulness-based approach – as the semi-structured focus-group interviews showed. One idea could be acquainting the whole school with the HAND in HAND approach by doing some exercises in small groups during a few training sessions and, consequently, offering extensive support and supervision for those interested.

A second central suggestion for improvement that can be derived from the formative evaluation results is realizing an encompassing whole-school approach. Whole-school change can be conceptualized to include changes in material conditions (e.g., infrastructure, resources, staff competencies), institutional processes (e.g. approaches to curriculum planning) and everyday practices in the school (in particular, changing the hidden curriculum²⁸) and to include top-down as well as bottom-up initiatives (e.g. Darling-Hammond, 2005; Downes, 2014; Ferreira, Ryan, & Tilbury, 2006; Thomson, 2010). Such a multi-layered change process is likely to require a much longer programme, as argued above. Additionally, it would help to include a critical mass of stakeholders in each school and to foster exchange between them. The HAND in HAND programmes addressed only the students in only one single class, the teachers teaching this class plus school leaders/other school staff and these three groups had separate trainings. It is not surprising that this is not yet sufficient for changing school cultures. A related issue is better involving the specialized school staff, who already have a focus on the quality of social relations in school such as school social worker, school psychologists, school counsellors and special needs teachers, in the planning and implementation of the programme – an aspect also raised by many of the adult participants during semi-structured focus-group interviews. More fundamentally, a participatory approach that involves teachers and other school staff in goal setting, planning and design of the school improvement process might be advisable (Hopkins, 2005). Also integrating the HAND in HAND programmes with existing school improvement strategies might help increasing sustainability of change (ibid.). Finally, single participants suggested integrating social learning with the curricula and getting compensatory time-off, which raises the point of changing infrastructure and institutional processes in order to support teachers with changing their pedagogical practice, which is also related to the idea of a “whole-school approach”.

²⁸ »[lessons] which are learned but not openly intended« such as norms, values, and beliefs transmitted in schools, see e.g. Martin (1983).

Noticeable is also the difference between stakeholders in their evaluation of the programmes, i.e. that students' evaluations were more critical than those of adults. Also, both students and trainers reported discipline problems during the student programmes suggesting a lack of motivation on the side of students. This might be related to the choice of the target age group: A meta-analysis by Yeager, Fong, Lee and Espelage (2015) suggested that anti-bullying programmes are less effective with students in grade eight or older than with younger students. Accordingly, school staff argued during the interviews that addressing younger students would be beneficial. This might be one way to go for future trials of the HAND in HAND programme. Besides, increasing autonomy support for students during the programme could be helpful. Downes and Cefai (2016) argued: "With older students, the question also arises as to their particular resistance to didactic style approaches that would undermine their increased sense of autonomy" (p. 39). In contrast to this desire for autonomy in the target group, none of the schools involved students in the decision to participate in the HAND in HAND programme in the first place, as the semi-structured focus-group interviews suggest. Also, the programme itself offered little choices to students – adaptations to the local contexts were made by the trainers, not by the participants and not even in consultation with the participants. Allowing for more student participation in decision-making at different stages of programme implementation might increase students' motivation to participate (see e.g. Deci & Ryan, 1987; Ryan & Deci, 2000; for a further discussion of autonomy support strategies see e.g. Assor, Kaplan, & Roth, 2002; Deci & Ryan, 1987; Ryan & Deci, 2000; Stefanou, Perencevich, DiCintio, & Turner, 2004). Otherwise, the interviews also showed that students were particularly interested in reflecting diversity, which suggests that increasing the share of exercises addressing diversity in the programme for students might also help better reaching out to the students (at least when the same age group would be addressed again).

Apart from the exercises addressing diversity – which many student groups commented positively on and only few participant criticized – the HAND in HAND exercises received varied feedback from participants. Some participants liked some exercises and others liked other exercises better. Thus, it seems advisable to keep a good mix so that there is something for everybody's taste. However, some issues with specific exercises were repeatedly mentioned. First of all, many of the adult participants did not like the exercises involving dialogue

and it seems that many had not completely understood the method and found it too demanding. Here, more modelling by trainers and more feedback from trainers might be necessary. Many students further described challenges with working in groups. In classes with little experience with cooperative learning it might, thus, be helpful to introduce and practice norms for group work more specifically and to take some time to reflect problems and conflicts during group work afterwards and with the whole class. Of course the HAND in HAND training itself addresses competencies needed for successful group work, but it might help to additionally teach specific strategies for addressing problems that are typical for cooperative learning (see e.g., Cohen & Lotan, 2014).

Apart from these larger topics, participants also made some specific suggestions for small changes that might be helpful for improving the programmes. The main points were: Explaining better the purpose of exercises, giving handouts, doing the training with smaller groups and using a more simple language during the training.

Hence, all in all, the formative evaluation of the HAND in HAND programmes helps understanding possible reasons for the lack of effects of the HAND in HAND programme and provide manifold ideas for improving the programmes. Suggestions for improvement of the HAND in HAND programmes that can be derived from the formative evaluations are summarized in the following:

1. Extending the programmes:
 - Add more training sessions; have training sessions over a whole school year
 - Add, in particular, content and methods that have the aim to help participants implement new skills in everyday interactions/pedagogical practice
2. Realizing a full “whole-school approach”:
 - Provide a programme for more stakeholders within each school, preferably for all students and for all members of the school staff
 - Address not only individual competencies and practices, but also material conditions in the school and institutional processes, such as curriculum planning, and also support related bottom-up initiatives.
3. Realizing a more participatory approach/ better support participants’ autonomy to increase motivation
 - Involve all participants in the decision to participate in the programme in the first place
 - Involve all participants in goal setting, planning and design of the school improvement process
4. Revision of some exercises
 - Most exercises appear to have worked well and participants interview responses suggest that there is a good mix of exercises
 - Often criticized was the exercise involving dialogue. This might need some revision (maybe a more detailed instructions and more common practice, because many participants seem not to have fully understood the exercise)
 - The chair exercise was often criticized for being potentially dangerous.
 - Students liked the exercises addressing diversity in particular - maybe some more of this type could be included in the programme
 - As several student complained about difficulties during group-work, adding a skill-builder for cooperative-learning (e.g., Cohen & Lotan, 2014) in advance to implementing group discussion over sensitive topics might be helpful – at least when the programme is applied in classes where students are not so familiar with cooperative learning.

12.4. Conclusions

The HAND in HAND programmes have been a positive experience for many of the participants and also for the trainers. Many participants found the programmes interesting and useful. Also, there is evidence that trainers often succeeded in creating a positive atmosphere and that some of the adult participants bonded as a result of their participation. Yet, the programmes have not caused consistent and stable changes neither in participants' social, emotional and/or inter/transcultural competencies nor in the social climates of participating classes (although there are some indications that the ability to observe unjudgementally of many participants was improved by the programmes). Possible reasons for this lack of effects – brought up by participants themselves – are that the programmes were too short to support change in behaviours, routines and practices and that only a small group of students, teachers and school staff in each school was addressed. Hence, it might be worthwhile to examine in future studies whether more extensive programmes that address not only single classes and a selection of the teaching body in a school, but really whole schools as dynamic systems will be more effective.

12.5. References

- Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours in predicting student's engagement in school work. *British Journal of Educational Psychology*, 72, 261-278.
- Bowman, C. L., & McCormick, S. (2000). Comparison of peer coaching versus traditional supervision effects. *Journal of Educational Research*, 93, 256–262
- Cohen, E. G., & Lotan, R. A. (2014). *Designing groupwork: Strategies for the heterogeneous classroom*. New York: Teachers College Press.
- Collaborative for Academic, Social, and Emotional Learning. (2003). *Safe and sound: An educational leader's guide to evidence-based social and emotional learning (SEL) programmes*. Chicago, IL: Author.

- Darling-Hammond, L. (2005). Policy and change: Getting beyond bureaucracy. In A. Hargreaves (Eds), *Extending educational change* (pp. 362-387). Dordrecht: Springer.
- Deci, E. L. & Ryan, R. M. (1987). The support of autonomy and the control of behaviour. *Journal of Personality and Social Psychology*, 56(6), 1024–1037.
- Downes, P. (2014). Conceptual Framework and Agenda: Beyond Bronfenbrenner (1979, 1995) to Interrogation of Blocked Systems via Structural Indicators. In P. Downes (Eds). *Access to Education in Europe: A Framework and Agenda for System Change* (pp. 29-45). Dordrecht: Springer.
- Downes P. & Cefai, C. (2016). *How to Prevent and Tackle Bullying and School Violence: Evidence and Practices for Strategies for Inclusive and Safe Schools, NESET II report*. Luxembourg: Publications Office of the European Union.
- Ferreira, J., Ryan, I., & Tilbury, D. (2006). *Whole-school approaches to sustainability: a review of models for professional development in pre-service teacher*. Sydney: Australian Research Institute of Education for Sustainability for the Australian Government Department of the Environment and Heritage. URL: <http://aries.mq.edu.au/projects/preservice/files/TeacherEduDec06.pdf>
- Glatthorn, A. A. (1987). Cooperative professional development: Peer-centered options for teacher growth. *Educational Leadership*, 45(3), 31–35.
- Guskey, T. R. (2000). *Evaluating Professional Development*. Thousand Oaks, CA: Corwin Press.
- Hargreaves, A., & Dawe, R. (1990). Paths of professional development: Contrived collegiality, collaborative culture, and the case of peer coaching. *Teaching and Teacher Education*, 6(3), 227-241.
- Hopkins, D. (2005). *The practice and theory of school improvement: International handbook of educational change*. Dordrecht, Netherlands: Springer.
- Naumann, A., Rieser, S., Musow, S., Hochweber, J., Hartig, J. (2019). Sensitivity of test items to teaching quality. *Learning and Instruction*, 60, 41-53.

- Naumann, A., Hartig, J., Hochweber, J. (2017). Absolute and relative measures of instructional sensitivity. *Journal of Educational and Behavioral Statistics*, 42(6), 678-705.
- Naumann, A., Hochweber, J., Klieme, E., (2016). A psychometric framework for the evaluation of instructional sensitivity. *Educational Assessment*, 21(2), 89-101
- Ryan, R. & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 78–88.
- Showers, B. (1984). *Peer Coaching: A Strategy for facilitating transfer of programme*. Eugene, OR: Center for Educational Policy and Management.
- Ryan, R. & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 78–88.
- Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychologist*, 39(2), 97-110.
- Thomson, P. (2010). *Whole school change: a literature review* (2nd edition). Newcastle: Creativity, Culture and Education. URL: <http://rummelighed.org/wp-content/uploads/2015/10/cce-lit-review-whole-school-change-255.pdf>
- Yeager, D. S., Fong, C. J. M., Lee, H. Y., Espelage, D. L. (2015). Declines in efficacy of anti-bullying programmes among older adolescents: Theory and a three-level meta-analysis. *Journal of Applied Developmental Psychology*, 37(1), 36-51.