

## Appendix 1

This appendix contains the item texts in Danish and English as well as the detailed results of the item analyses by Rasch models.

Table A1. *The items in the Critical Thinking Scale*

Items	Danish (used in study)	English equivalent
CTh1	Jeg stiller tit spørgsmålstegn ved ting, jeg hører eller læser i statistik-faget, for at finde ud af, om jeg synes, de virker overbevisende.	I often find myself questioning things I hear or read in this statistics course to decide if I find them convincing
CTh2	Når en teori, en fortolkning eller en konklusion præsenteres i statistik-faget, forsøger jeg at afgøre, om den er velunderbygget	When a theory, interpretation or conclusion is presented in the statistics course or in the readings, I try to decide if there is good supporting evidence
CTh5	Når jeg hører eller læser en påstand i statistik-faget, overvejer jeg alternative forklaringer.	Whenever I read or hear an assertion or conclusion in this statistics course, I think about possible alternatives

Notes. Item numbers correspond to the numbers used in Nielsen et al. (2021).

Table A2. *Conditional likelihood ratio tests of local independence of item at baseline and follow up*

item pairs tested	Baseline			Follow-up		
	CLR	df	p	CLR	df	p
CTh1 & CTh2	14.50	16	0.561	8.66	16	0.927
CTh1 & CTh5	34.44	16	0.005 <sup>+</sup>	8.52	16	0.932
CTh2 & CTh5	15.53	16	0.486	13.12	16	0.664

Notes. Benjamini-Hochberg correction for FDR rejects at 0.00278 at the 5% level

Table A3. *Conditional likelihood ratio tests of no DIF relative to five background variables at baseline and follow-up*

Items within time points	<u>Year cohort</u>			<u>Math adequacy</u>			<u>Background variables</u> <u>Stat in future work</u>			<u>Gender</u>			<u>Age groups</u>		
	CLR	df	p	CLR	df	p	CLR	df	p	CLR	df	p	CLR	df	p
Baseline															
CTh1	3.26	4	0.515	4.64	4	0.326	16.74	8	0.033 <sup>+</sup>	7.21	4	0.125	3.81	4	0.433
CTh2	1.80	4	0.772	0.85	4	0.931	10.08	8	0.259	4.17	4	0.384	6.03	4	0.197
CTh5	9.19	4	0.057	5.56	4	0.234	7.95	8	0.438	3.75	4	0.441	6.82	4	0.145
Follow-up															
CTh1	2.31	4	0.679	0.56	4	0.968	6.67	8	0.573	1.10	4	0.894	2.10	4	0.718
CTh2	5.79	4	0.216	3.74	4	0.443	3.54	8	0.896	1.38	4	0.848	2.92	4	0.571
CTh5	2.11	4	0.716	4.14	4	0.387	19.64	8	0.011 <sup>+</sup>	5.01	4	0.287	1.36	4	0.851

Notes. Benjamini-Hochberg correction for FDR rejects at 0.00278 at the 5% level

Table A4. *Item thresholds, locations, difficulties, targets and information at baseline and follow-up*

items at time points	<u>Thresholds</u>				Location <sup>a</sup>	Item difficulty <sup>b</sup>	Item target <sup>c</sup>	Information at target
	1	2	3	4				
Baseline								
CTh1	-3.51	-0.64	1.05	3.78	0.17	0.20	0.25	0.50
CTh2	-2.96	-0.75	0.23	2.52	-0.24	-0.26	-0.29	0.67
CTh5	-1.90	-0.50	0.41	2.27	0.07	-0.01	-0.22	0.79
Follow-up								
CTh1	-3.33	-0.88	1.19	4.41	0.35	0.17	-0.26	0.45
CTh2	-3.15	-1.35	0.41	4.00	-0.03	-0.42	-1.27	0.54
CTh5	-2.62	-1.21	-0.05	2.60	-0.32	-0.56	-1.06	0.70

Notes.

<sup>a</sup>. Mean of the thresholds

<sup>b</sup>. the person parameter where item information is maximized.

<sup>c</sup>. the person parameter where the expected score = max score/2.

Table A5. *Targeting and reliability of the Critical Thinking Scale at baseline and follow-up*

time point	target	mean	TI mean	TI max	Theta	RMSE mean	RMSE min	RMSE target index	target	Score		r
					TI target index					mean	mean SEM	
Baseline	-0.17	-0.78	1.616	1.954	0.827	0.801	0.715	0.893	8.75	7.78	1.26	0.72
Follow-up	-1.01	-0.76	1.447	1.678	0.862	0.838	0.772	0.921	7.86	8.26	1.20	0.75

*Notes.* TI = test information, RMSE = The root mean squared error of the estimated theta score. SEM = The standard error of measurement of the observed score. r = reliability

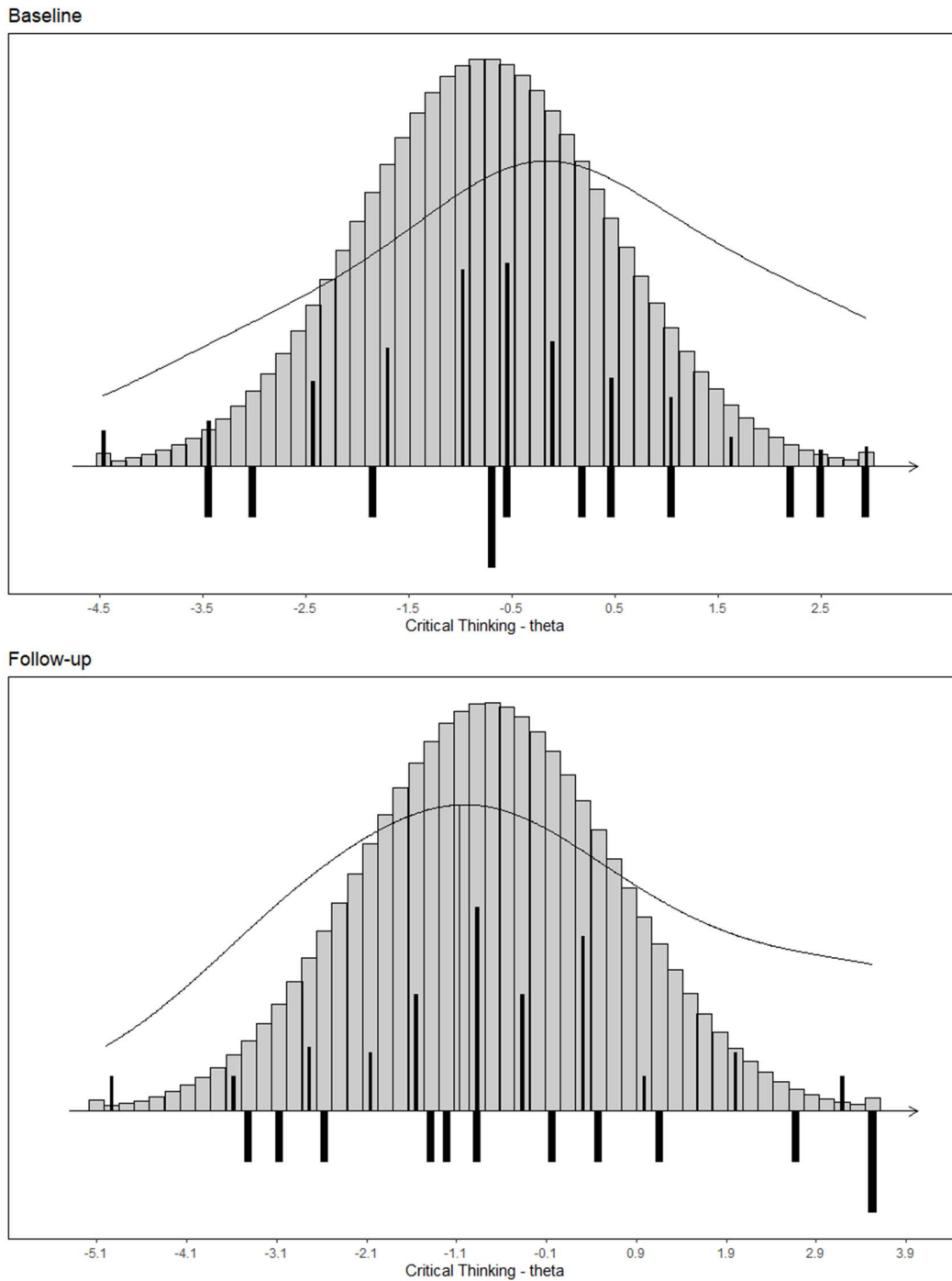


Figure A1. Item maps showing distribution of person parameter estimates and information curve above item threshold locations for the Critical Thinking scale at baseline and follow-up

Notes. Person parameters are weighted maximum likelihood estimates and illustrate the distribution of these for the study sample (black bars above the line) and for the population under the assumption of normality (grey bars above the line), as well as the information curve, relative to the distribution of the item thresholds (black bars below the line).

Table A6. Conversion of summed scale scores to weighted maximum likelihood estimates of person parameters and rescaled person parameters for the Critical Thinking Scale at baseline and follow-up

Summed scale score	<u>Baseline</u>		<u>Follow-up</u>	
	Person parameter estimates	Rescaled person parameter estimates	Person parameter estimates	Rescaled person parameter estimates
3.00	-4.922	3.00	-4.977	3.00
4.00	-3.405	4.83	-3.599	4.52
5.00	-2.413	6.03	-2.772	5.44
6.00	-1.638	6.96	-2.095	6.19
7.00	-1.032	7.69	-1.492	6.85
8.00	-0.523	8.31	-0.926	7.48
9.00	-0.049	8.88	-0.354	8.11
10.00	0.439	9.47	0.271	8.80
11.00	0.992	10.13	1.021	9.63
12.00	1.655	10.93	1.987	10.70
13.00	2.457	11.90	3.152	11.99
14.00	3.455	13.10	4.324	13.29
15.00	5.028	15.00	5.874	15.00

*Notes.* Summed scale scores are raw scores. Person parameters are interval-scaled weighted maximum likelihood estimates of the person parameters resulting from the Rasch models. Rescaled person parameters are the WML person parameters rescaled to the original range of the summed raw score, but now with equal distance between values

## Appendix 2

Appendix 2 contains the relevant extracts from the curriculum documents for the bachelor of psychology degree and the statistics course the participants were enrolled in are presented below with brief explanations<sup>1</sup>.

Highlights were made by the authors to show where we find the curriculum to reflect critical thinking either explicitly or implicitly, and thus these are only expressions of our perspective.

### End competencies for bachelor of psychology

Most Danish university degree programs describe the objectives which should be the result of the full degree in terms of the knowledge, skills and competencies students should master.

Upon graduation, the bachelor in psychology should be able to:

#### Knowledge

- have a broad research-based knowledge of and insight into the theories and methods of psychology.
- have a basic understanding of and insight into concepts, theories and methods in personality, cognitive, developmental and social psychology.
- have knowledge of and be able to reflect on psychological conditions concerning organizational and work psychology, clinical and educational psychology.

#### Skills

- be able to apply the scientific methods and tools of psychology, identify theoretical and practical issues within the field of psychology.
- be able to identify and illuminate general human conditions that involve ethical, cultural and other conflicts.
- be able to select and justify relevant analysis and solution models in relation to psychological topics and tasks.
- be able to convey general psychological issues to peers and non-peers. That is: explain how a psychological topic is treated within different psychological disciplines and based on different theoretical assumptions and their methodological and empirical basis.

#### Competencies

- be able to handle psychological tasks of less complexity where psychological knowledge is required, such as
  - o cases concerning individuals, groups and organizations,
  - o establishment and maintenance of customer / client contact
  - o participation in projects that create better conditions for children, young people, adults and the elderly,
  - o participation in psychological development projects
  - o assist in psychological research projects

---

<sup>1</sup> All documents are readily available online and the text in the boxes are quoted directly.

- be able to independently intervene in situations that require psychological insight by analyzing, structuring and setting goals within a practical or theoretical area, and thus plan, implement and evaluate courses of action in collaboration with people in the psychological field as well as from other fields. The development of these collaborative relationships also involves collaboration between students.
- be able to identify own learning needs and structure own learning processes in connection with solving the above tasks

### Course learning objectives

Course learning objectives are also described in terms of the knowledge, skills and competencies students should master at the end of the course or part of course.

After the first semester of the statistics course the student should be able to:

#### Knowledge

- Demonstrate knowledge of the most central terminology and their meaning in statistics.
- Account for and understand statistical results.
- Know the basic possibilities and operating principles in a statistical software, e.g. SPSS.

#### Skills

- Operate analytically when choosing statistical methods.
- Plan and execute statistical analyses of simple studies, ex with two groups.
- Execute statistical analyses by use of a statistical software, e.g. SPSS.
- Mediate statistical conclusions with correct terminology.

#### Competencies

- Use suitable statistical methods to solve psychological problematics.
- Draw conclusions about unknown sizes and parameters attached to the population.

### Assessment criteria

Student performance are assessed within a locally developed model with four levels of generic competencies. The top level covers the PhD level (level 1), while levels four to two cover bachelor and master programs. The model is constructed so that lower levels are also included in higher levels. As the statistics course in question is described as a level 3 course, both the lower level 4 and the level 3 descriptions are shown below (*translated from Danish*):

levels	Scientific ability	Practical skills
level 3:	To account for a psychological field or a psychological problem and analyse field- or problem-related relationships between	To apply given, scientifically justified methodologies and forms of practice in relation to concrete problems.

	relevant theories, methods and empirical knowledge.	
level 4:	To account for basic features of a psychological field or important features of a psychological problem by including relevant theories, methods and empirical knowledge.	To apply given, scientifically justified methodologies and forms of practice to defined tasks.

The assessment criteria are described for all courses in terms of what is needed to obtain the highest possible grade (i.e. 12) and what is needed to obtain the lowest possible passing grade (i.e. 02). When a course exam is not graded using the Danish grading scale (included at the end of this appendix), the descriptions are provided in the same manner, but then describes the best possible pass and the worst possible pass. Anything below the latter will be a fail. Descriptions are translated from Danish.

<p>Performance worthy of 12:</p> <p><i>Excellent</i>, which means:</p> <p>The performance – based on an overall assessment in an exhaustive manner and without significant shortcomings – fulfils the course goals, which are to:</p>	<p>Performance to 02:</p> <p><i>Adequate</i>, that is:</p> <p>The performance – based on an overall assessment to a sufficient extent and minimally acceptable degree – fulfils the course goals, which means that:</p>
<p>Carry out simple statistical studies, including reformulating professional problems as statistical problems; choosing appropriate statistical methods to solving the problem; interpreting the results and translating statistical conclusions into non-specialist terms.</p>	<p>The formulation of the statistical problems based on professional problems, the choice of statistical method and the interpretation of the results are inadequate.</p>
<p>Read scientific articles that are based partly on a statistical study and on being able to respond critically to the choice of method as well as the conclusions that are drawn on the basis of the statistical material.</p>	<p>The explanation of the choice of method and conclusions are inadequate or completely disregarded.</p>
<p>Be able to critically evaluate statements using statistics or where statistics have been used.</p>	<p>The critical assessment of statements using statistics is unclear and inadequate.</p>



**The Danish grading scale<sup>a</sup>**

Grade	Requirement for awarding grade
12	For an excellent performance displaying a high level of command of all aspects of the relevant material, with no or only a few minor weaknesses.
10	For a very good performance displaying a high level of command of most aspects of the relevant material, with only minor weaknesses.
7	For a good performance displaying good command of the relevant material but also some weaknesses.
4	For a fair performance displaying some command of the relevant material but also some major weaknesses.
02	For a performance meeting only the minimum requirements for acceptance.
00	For a performance which does not meet the minimum requirements for acceptance.
-3	For a performance which is unacceptable in all aspects.

*Notes.* Grades 00 and -3 are failing grades, the remainder are passing grades

<sup>a</sup> official translation from the Ministry of Higher Education and Science