Chapter 4
The Heidelberg Inventory of Geographic System Competency Model

Kathrin Viehrig, Alexander Siegmund, Joachim Funke, Sascha Wüstenberg, and Samuel Greiff

Abstract The concept "system" is fundamental to many disciplines. It has an especially prominent place in geography education, in which additionally, the spatial perspective is central. Empirically validated competency models dealing specifically with geographic systems—as well as adequate measurement instruments—are still lacking. Therefore, based on the theoretically-guided development of a Geographic System Competency (GSC) model, the aim was to build and evaluate such a measurement instrument, with the help of probabilistic measurement models. The competency model had three dimensions: (1) "comprehend and analyze systems", (2) "act towards systems" and (3) "spatial thinking", whereby dimension (2) was changed to "evaluating possibilities to act towards systems" after a thinking-aloud study. A Cognitive Lab (CogLab) and two quantitative studies (Q1 $n = 110$, Q2 $n = 324$) showed divergent results. Dimension (2) could not be identified in both quantitative studies. Whereas Dimensions (1) and (3) constituted separate dimensions in Q1, in Q2 the two-dimensional model did not fit significantly better than the one-dimensional model. Besides showing the close relationship between spatial and systemic thinking in geographic contexts, which are thus both needed in modeling GSC, the project highlights the need for more research in this central area of geography education.

K. Viehrig (✉)
School of Education, University of Applied Sciences and Arts Northwestern Switzerland, Windisch, Switzerland
e-mail: kathrin.viehrig@fhnw.ch

A. Siegmund
Heidelberg University of Education, Heidelberg, Germany
e-mail: siegmund@ph-heidelberg.de

J. Funke
Heidelberg University, Heidelberg, Germany
e-mail: funke@uni-hd.de

S. Wüstenberg • S. Greiff
University of Luxembourg, Esch-sur-Alzette, Luxembourg
e-mail: sascha.wuestenberg@uni.lu; samuel.greiff@uni.lu

D. Leutner et al. (eds.), *Competence Assessment in Education*, Methodology of Educational Measurement and Assessment, DOI 10.1007/978-3-319-50030-0_4

Keywords Systems thinking • Spatial thinking • Competence model • University students • Geography

4.1 The Role of Geographic System Competency in Geography Education

Geography often deals with complex human-environment systems that are seen as important to society and business. Whether it is a matter of extreme weather events, transformations in the energy sector or resource conflicts, learning to understand (geographic) systems has been a central part of the overall objective of geographic education for decades (e.g., DGfG 2010; Köck 1993).

Because the concept "system" is regarded as one of the most important cognitive constructs of science (e.g., Klaus 1985; Smithson et al. 2002), research looking at how learners and/or experts understand systems is undertaken in different subjects (e.g., physics: Bell 2004; mathematics: Ossimitz 2000; biology: Sommer 2005; economics: Sweeney and Sterman 2000), in interdisciplinary areas such as education for sustainable development (e.g., Rieß and Mischo 2008) and in the area of complex problem solving (e.g., Funke 1990). Thus, the research spans a wide age range from kindergarten to adults/university.

Geographic inquiry deals with "[…] the whys of where […]" (Kerski 2013, 11). Consequently, to understand geographic systems, both systemic (why) and spatial thinking skills (where) seem necessary.

In general, systemic and spatial thinking would appear to be researched mostly independently of each other. Moreover, despite their longstanding importance in the German geography education discourse, the specific geographic competencies necessary to understand *geographic* systems seem not to have been empirically identified as yet, especially with regard to the relationship of systemic and spatial thinking. Additionally, there seem to be only few validated, psychometrically and geographically adequate, assessment instruments.

Consequently, in recent years, two DFG-funded projects have started to test competency models empirically for geographic system competency. The model by Rempfler and Uphues (see e.g., 2010, 2011, 2012) is based on a socio-ecological approach and focuses on systemic thinking. In contrast, the Heidelberg Inventory of Geographic System Competency (HEIGIS) model explicitly includes both systemic and spatial thinking (Table 4.1).

Hence, in line with the general competency definition in the Priority Program (SPP 1293), and based on existing works (see overview in Viehrig et al. 2011; Viehrig et al. 2012), geographic system competency (GSC) has been defined as "[…] the cognitive achievement dispositions […] that are necessary to analyze, comprehend geographic systems in specific contexts and act adequately towards them" (Viehrig et al. 2011, p. 50, translated).

Table 4.1 Original HEIGIS model[a]

	Dimension 1: Comprehend and analyze systems	Dimension 2: Act towards systems	Dimension 3: Spatial thinking
Level 3	Identification and understanding of the complex network of relationships	Also take into account side effects and autoregressive processes	Use several spatial thinking skills in a structured way
Level 2	Identify and understand relationships between the system elements	Take into account multiple effects	Use several spatial thinking skills in an unstructured way
Level 1	Identify and understand system elements	Take into account main effects	Use only one spatial thinking skill in an unstructured way

Largely based on Ben-Zvi Assaraf and Orion (2005), Gersmehl and Gersmehl (2006), Greiff and Funke (2009), and Hammann et al. (2008)
[a]Viehrig et al. (2011, p. 51, translated)

The original model (Table 4.1) comprised three dimensions: *comprehend and analyze systems*, *act towards systems* and *spatial thinking*, with differentiation between Dimensions 1 and 2 based both on geographic education theory (e.g., Köck 1993, p. 18, translated, speaks of "thinking and acting in geo-ecosystems") and empirical results in problem solving (e.g., Greiff 2010). An overview of the basis for development can be found, for example, in Viehrig et al. (2011) and Viehrig et al. (2012). The spatial thinking skills in Dimension 3 refer to those identified by Gersmehl and Gersmehl (2006), namely: "Defining a Location […]" (p. 12), "Describing Conditions […]" (p. 13), "Tracing Spatial Connections […]" (p. 14), "Making a Spatial Comparison […]" (p. 14), "Inferring a Spatial Aura […]" (p. 15), "Delimiting a Region […]" (p. 15), "Fitting a Place into a Spatial Hierarchy […]" (p. 16), "Graphing a Spatial Transition […]" (p. 17), "Identifying a Spatial Analog […]" (p. 18), "Discerning Spatial Patterns […]" (p. 19) and "Assessing a Spatial Association […]" (p. 20) (partly bold-enhanced in the original).

To test the HEIGIS model empirically, three different studies were conducted. The studies were targeted at university students in subjects including geography, geography education and other subjects, such as psychology. However, the first one in particular was constructed with thought to its applicability to high school in mind.

4.2 Study Overview

The three studies conducted within the project (2009–2011) consisted of a videographed thinking-aloud cognitive lab study (CogLab) split into two rounds, and two quantitative studies (labeled Q1 and Q2). An overview over the samples used for analysis can be seen in Table 4.2.

The CogLab aimed at further developing the competency model as well as exploring possible similarities in domain-general problem solving. The CogLab

Table 4.2 Sample overview used for analysis[a]

	CogLab Round 1 & 2	Q1	Q2
n	10	110 (questionnaire)	324
		67 (MicroDYN)	
Students	100 %	98.2 %	96.6 %
Type of students	Pre-service teacher geography 90.0 %	Psychology 33.6 %	psychology 17.9 %
	Pre-service teacher other 10.0 %	Pre-service teacher geography 60.9 %	Pre-service teacher geography 37.0 %
		Geography 2.7 %	Geography 34.9 %
		Other 0.9 %	Other geo-sciences 4.9 %
			Pre-service teacher other 0.6 %
			Other 1.2 %
Male	50.0 %	32.7 % (1 missing)	39.2 %
M age (*SD*)	23.8 (2.3)	23.2 (7.2) (1 missing)	23.9 (6.1)
M GPA		2.1 (0.8) (8 missing)	2.3 (0.7)
GPA better than 2.5		55.5 % (8 missing)	53.4 %

GPA grade point average (school leaving certificate, with 1.0 considered the best and 4.0 considered passed)
[a]The description for Q1 refers to the 110 participants of the questionnaire

was conducted in two rounds. The first round ($n = 5$) used fictitious examples, while the second round ($n = 5$) used real world examples. Moreover, the two rounds differed in the items used.

Q1 aimed at a first quantitative exploration of the modified model and the relationship between GSC and problem solving. Thus, the study consisted of two parts: a questionnaire containing background variables, and including the geographic system competency items, and a MicroDYN to measure problem solving (for an introduction to the MicroDYN testing environment see the chapter by Funke and Greiff (2017, in this volume)). The GSC used real world examples. In Q1, 137 participants filled out the questionnaire, of which 110 were included in the analysis because they filled out at least one of the GSC items (only 81 returned complete questionnaires). In the MicroDYN part, there were 81 participants, of which 67 could be included in the analysis. The rest were excluded either because data was not saved properly or, for more than 25 % of the items, questions were not answered.

Q2 aimed at further exploring the structure of GSC, with the help of a revised questionnaire and a larger sample. In Q2, there were over 600 participants, of which 324 were included in the analysis. Excluded participants included those who returned incomplete answers, and those reporting a below-B1[1] language level in German.

[1] Using a simplified self-report scale (A1 to C2 plus an option for native speaker), based on the "Common European Framework of Reference for Languages" (CEFR, see e.g. Council of Europe 2001)

This chapter focuses on the results related to the structure of GSC. The relationship of achievement to various variables included in the studies will be reported elsewhere.

4.3 CogLabs

4.3.1 Description of the Measurement Instruments

The CogLabs contained different item formats, especially MicroDYN, concept maps and short answer tasks in the first round, and MicroDYN, multiple-choice and "add to a started concept map" tasks in the second round (see details in Viehrig et al. 2011; Viehrig et al. 2012). Based on the national educational standards (DGfG 2010), items were generated in three areas, that is: physical and human geography, and human-environment interactions.

Concept maps are frequently used to measure domain-specific systemic thinking, both in the geo-sciences (e.g., Ben-Zvi Assaraf and Orion 2005) and in other subjects, such as biology (e.g., Sommer 2005). Short answer questions are often used in educational courses and have also been used in systemic thinking research (e.g., Sommer 2005). MicroDYN items have been used to measure problem solving and have shown good psychometric properties. They consist of minimally complex systems, with students first having three minutes to find out the structure of the system and then one and a half minutes to achieve a specified aim by manipulating system variables (e.g., Greiff 2010). A brief discussion of the advantages and disadvantages of some item formats can be found, for example, in Viehrig et al. (2011). In both rounds the systems used were very simple, in order to fit with the minimally complex structure of MicroDYN.

The first CogLab round started with a general introduction by the respective interviewer, the signing of a consent form and a short questionnaire collecting basic demographic data. Then the measurement instrument proper began with an example and explanation of the MicroDYN format for the students to explore. Afterwards, the students had to respond to six MicroDYN items, to measure their problem solving skills. This part was followed by three MicroDYN items using geographic contexts to measure Dimension 1 (part: model building) and Dimension 2 (part: prognosis). The geographic system competency and the problem solving items had identical structures. The second part started with an example and explanation of CMapTools, software that can be used to create concept maps (available from http://cmap.ihmc.us/). The three tasks created consisted of a short informational text, a Dimension 1 item, which asked students to create a concept map, and a short answer item, approximating Dimension 2. This was followed by three tasks to measure Dimension 3. Students were presented with a number of simple thematic maps and had to use a concept map to describe their answers to a spatial question. Besides being asked to think aloud while responding to the items, there were specific questions, for example regarding their problems with a task, or what procedure they used to solve the task during the CogLab. Furthermore, there were some general questions after all tasks were completed: for example, what kind of similarities and differences they noticed in their thinking processes, between explicitly spatial and not explicitly spatial tasks. At the end, formalities regarding the participant's payment were taken care of. Sample items can be seen in Fig. 4.1.

The second CogLab round also started with a general introduction by the respective interviewer, the signing of a consent form and a short questionnaire collecting some basic demographic data. Afterwards, after a brief introduction, there were three tasks, with three items each for Dimension 1. After reading an informational text, the students had to answer two multiple-choice tasks (one correct answer) and

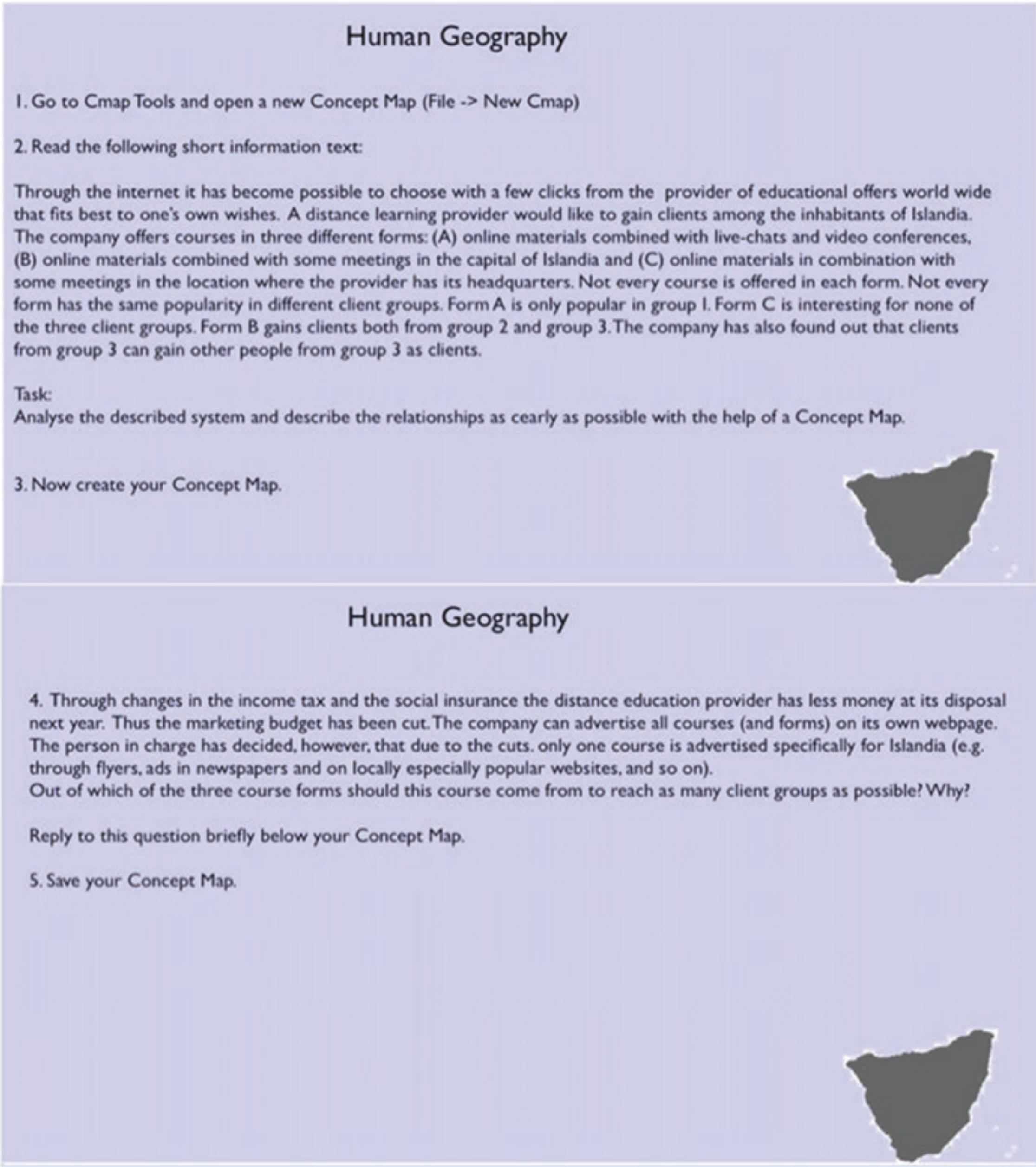

Fig. 4.1 Sample items from the CogLab Round 1: Concept map and short answer (Dimensions 1 and 2, translated)

Climate

Galveston is located on the coast of the Gulf of Mexico, in the South of the USA, in Texas.
In the media a lot is discussed about the global warming and its consequences. One possible consequence is the rise of the sea level[1].
Several factors can contribute to the rise of the sea level. This includes e.g. the expansion of the seawater through an increase in water temperature. Also the melting of the glaciers and ice caps on land can contribute to sea level rise.

Item 3: Complete the following graphic based on the text. Use only one of the following phrases per blank. For that write the suitable number in the blank.
(1) melting of ice on the land (e.g. glacier); (2) melting of ice bergs that swim on the water; (3) contributes to the rise of; (4) contributes to the lowering of the; (5) Gulf of Mexico Golf; (6) expansion of the seawater at warmer water temperature; (7) does not contribute to the rise of

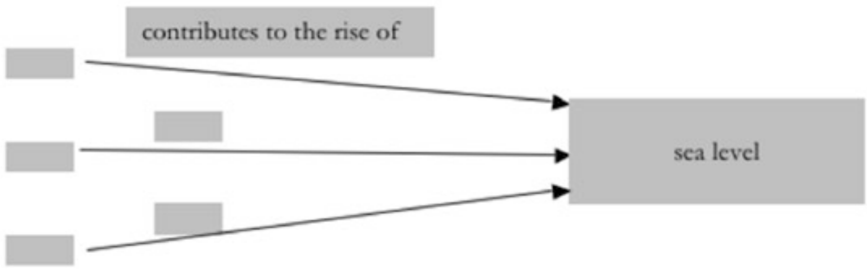

Fig. 4.2 Sample item from the CogLab Round 2: Item group stem and “add to a started concept map ” task (Dimension 1, translated)

one “add to a started concept map” task (concepts, relationship descriptions and structure provided). Then came the MicroDYN part, consisting of three geographic items to measure Dimensions 1 (part: model building) and 2 (part: prognosis). This was followed by three tasks, each consisting of a short informational text and two maps with three multiple-choice items (one correct answer), to measure the students’ spatial thinking skills. Because these items dealt with real-world examples, the students had to locate 19 countries on a world map and indicate that two were not a country, as a basic indicator of geographic pre-knowledge (spatial framework of reference). The last part consisted of four MicroDYN items to measure problem solving. The CogLab ended with some general questions and again, taking care of the formalities. A sample item can be seen in Fig. 4.2.

4.3.2 Selected Results

In the first CogLab, fictitious places were used as examples to reduce the influence of pre-acquaintance with a specific location (similar to e.g., Ossimitz 1996). In the second round, real places were used as examples. In summary, the CogLabs indicated a better suitability of real world examples to assess GSC, in terms of item generation and processing. Moreover, some of the participants emphasized the importance of the authenticity/real world relevance of the tasks. Participants also

Table 4.3 HEIGIS model after the CogLabs (translated)

	Dimension 1: Comprehend and analyze systems	Dimension 2: Evaluate possibilities to act towards systems	Dimension 3: Spatial thinking
Level 3	Identification and understanding of the complex network of relationships	Also take into account side effects and autoregressive processes	Can use 8 or more spatial thinking skills
Level 2	Identify and understand relationships between the system elements	Take into account multiple effects	Can use 6 or 7 spatial thinking skills
Level 1	Identify and understand system elements	Take into account main effects	Can use up to 5 spatial thinking skills

Largely based on Ben-Zvi Assaraf and Orion (2005), Gersmehl and Gersmehl (2006), Greiff and Funke (2009), and CogLab results

indicated, however, that the country example used in a task makes a difference, even if the essential information is given.

In general, the CogLabs hinted at the separability of Dimensions 1 and 3. However, the CogLabs resulted in two key changes to the competency model (see Table 4.3 and Viehrig et al. 2012). Firstly, the CogLabs indicated that MicroDYN is well suited to measuring general dynamic problem solving but not to content- specific systemic thinking. One example is Participant 9, who stated for instance (excerpt, translated):

> […] I really don't need my pre-knowledge and I also don't need to think about anything. […] So, I can simply try out. Then I see what happens there. That means I don't need to supply pre-considerations. I can try out as much as I want. That means I'm not forced at all to think for myself. […] Because I can, well, execute. I can simply look, how, with what controller moves […] the value. That means, I don't need to at all, eh, that could also be Chinese now, which I don't understand. I still would know which value has an effect on it. And I wouldn't know what's behind it. […] Without that I've understood it. Solely through the technicality. I put the controller up and that one changes and this one doesn't change, thus it's that one […].

Without MicroDYN, acting towards systems was no longer possible as an item format. This made it necessary to change Dimension 2 to proximal estimation, via the evaluation of several possibilities for acting towards systems.

Secondly, the CogLabs indicated that the levels of spatial thinking (assumptions based on Hammann et al. 2008) could not be observed in the concept maps. Moreover, while multiple choice items seemed to work for Dimension 1, the levels of Dimension 3 would hardly be representable by multiple choice tasks. Consequently, the levels were replaced by a preliminary quantitative graduation: that is, how many spatial thinking skills could be used, based on rounded 50 % and 75 % cutoffs. Gersmehl and Gersmehl (2007) state that "[t]he human brain appears to have several 'regions' that are structured to do different kinds of spatial thinking" (p. 181) and that "[p]arallel research by child psychologists and educational specialists tends to reinforce one main conclusion of the neuroscientists: the brain areas that are devoted to different kinds of spatial thinking seem to develop in very early childhood" (p. 188). Thus, there was no basis for assuming a systematic ranking in difficulty of the spatial thinking skills that could have been used for a more qualitative description of the levels.

4.4 First Quantitative Study (Q1)

4.4.1 Description of the Measurement Instruments

The first quantitative study (Q1) comprised two parts, namely, MicroDYN and a limesurvey questionnaire. The MicroDYN part consisted of six items and used geographic contexts. The limesurvey questionnaire contained background variables, eight items measuring interest in various aspects of geography on a five-point scale, as well as a geographic pre-knowledge task asking students to write the names of twelve countries marked on a world map, and three geographic pre-knowledge items asking students to choose one of five places (or none) where they would expect a specified condition to be fulfilled. This was followed by four GSC tasks for Dimensions 1–2 and six tasks for Dimension 3. Thus, the tasks contained 12 items to measure Dimension 1, 4 items to measure Dimension 2, and 14 items to measure Dimension 3. To get a credit, respondents had to check several correct and zero incorrect answers, bring answers into the correct sequence, etc. At the end, the participants were asked for their feedback, both in an open comment field and with the help of specific questions that they had to rate on a four point scale (e.g., regarding how much reading literacy the questionnaire requires). A sample item for the limesurvey questionnaire can be seen in Fig. 4.3.

4.4.2 Dimensions of the Competency Model

Firstly, each of the dimensions was tested separately for one-dimensionality using a CFA (confirmatory factor analysis) in Mplus (Muthén and Muthén 2007; Table 4.4) and a Rasch Analysis in Conquest (Wu et al. 2007). Thereby, for the analysis in Mplus, items with a small factor loading ($r_{it} < 0.40$) were excluded. The analysis in Mplus showed a good model fit in Dimension 1. There was a bad model fit for Dimension 3, which might have been caused by having only one item for each spatial thinking skill (except for "condition"; see overview of all skills in Gersmehl and Gersmehl 2006), due to test time considerations (see Table 4.4).

Because Dimension 2 was not identifiable/did not converge, it had to be excluded in further analyses of the data. The Rasch Analysis of the remaining items in Conquest showed acceptable WMNSQ (weighted mean square)- and *t*-values.

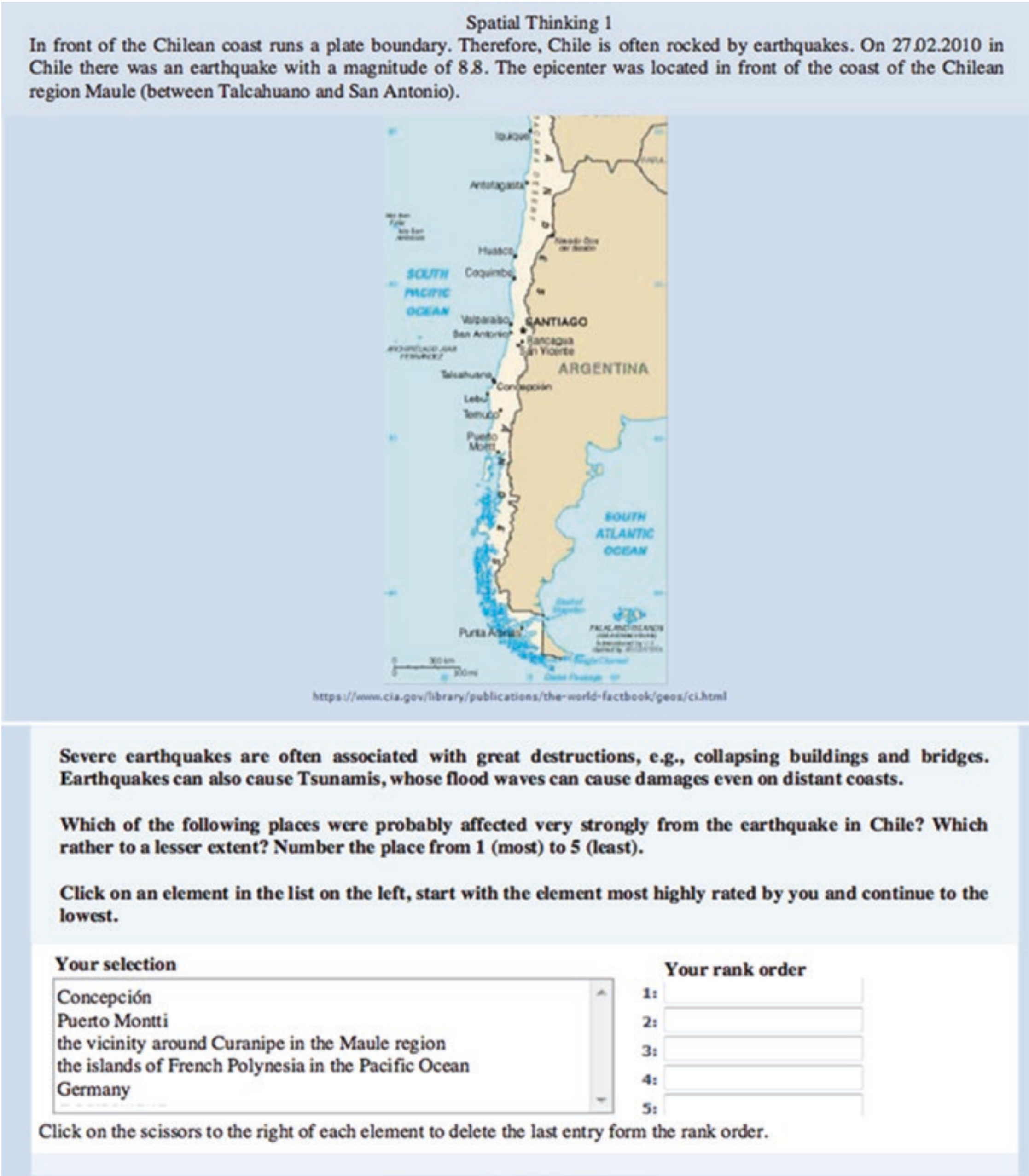

Fig. 4.3 Sample item from Q1: Limesurvey—Spatial thinking item stem and one of the associated items, namely, the one for the skill "Aura" (translated)

Secondly, a two- and a one-dimensional model were tested for the remaining dimension 1 and 3 items. As assumed, the two-dimensional model was to be preferred, both based on an analysis in Mplus and an IRT analysis in Conquest (Table 4.5) The χ^2 -test of difference for the Mplus analysis was calculated according to Muthén and Muthén (2007). The separability of the two dimensions is supported by a latent correlation of $r = 0.776$. This is fairly high. However, in PISA for example, even constructs with latent correlations >0.90 have been seen as separable (see Klieme et al. 2005).

Table 4.4 Results of the separate tests for one-dimensionality for each dimension in Q1 (Mplus)

	Dimension 1: Comprehend and analyze systems	Dimension 2: Evaluate possibilities to act towards systems	Dimension 3: Spatial thinking
Number of items	12	4	14
Number of items without excluded items	8	–	12
χ^2	15.138		42.093
df	13		26
p	>0.10		0.024
CFI	.98		.87
RMSEA	.04		.08
Conclusion	Remaining items fit one-dimensional model well	Model did not converge with all 4 items; model was not identifiable when excluding items	remaining items fit one-dimensional model barely acceptable

CFI Comparative Fit Index, *RMSEA* Root Mean Square Error of Approximation

Table 4.5 Results of the test for dimensionality (Dimensions 1 and 3) for the remaining items in Q1

Mplus	Two-dimensional model	One-dimensional model	χ^2 test of difference	
χ^2	53.727	56.915	χ^2	7.814
df	41	41	*df*	1
p	0.088	0.050	*p*	0.005
CFI	.92	.90		
RMSEA	.05	.06		
Conclusion	Two-dimensional to be preferred			
Conquest	Two-dimensional model	One-dimensional model	χ^2 test of difference	
Final deviance	1932.69	1945.51	χ^2	12.82
Number of estimated parameters	23	21	*df*	2
			p	0.002
Conclusion	Two-dimensional to be preferred			

4.4.3 Levels of the Competency Model

The GSC levels could only be examined on the basis of the remaining items of Dimensions 1 and 3. The item map, showing the difficulty of the items based on a two-dimensional Rasch model, can be seen in Fig. 4.5. "Condition" was the only spatial thinking skill for which more than one item was included, with "Condition 1" being assumed the easiest and "Condition 4" the hardest.

For Dimension 1, levels could not be confirmed, with the Level 1 items being unexpectedly difficult. There are several possible explanations (see also discussion in Viehrig 2015). Firstly, this could have been caused by the items used. Secondly, the empirically derived levels of Ben-Zvi Assaraf and Orion (2005) could possibly hold only in the context of recall tasks, and not in tasks in which information is provided in the item stem. Thirdly, item difficulty might be influenced by differences in terms of the sample's (German vs. Israeli, university vs. school students) prior educational experiences.

For Dimension 3, an analysis of the raw data sum score of the remaining Level 3 items indicated the possibility for using the number of spatial thinking skills as a graduation. Thereby, "condition" was counted if any of the tasks were solved. It must be kept in mind that one spatial thinking skill had to be excluded and that only 81 respondents could be included in the analysis. Level 1 was reached by 76.5 % of the sample, Level 2 by 18.5 % and Level 3 by the remaining 5 %. The item map (Fig. 4.4) shows differences in difficulty between the spatial thinking skills items. Moreover, the spatial thinking skill "condition" showed a graduation in difficulty

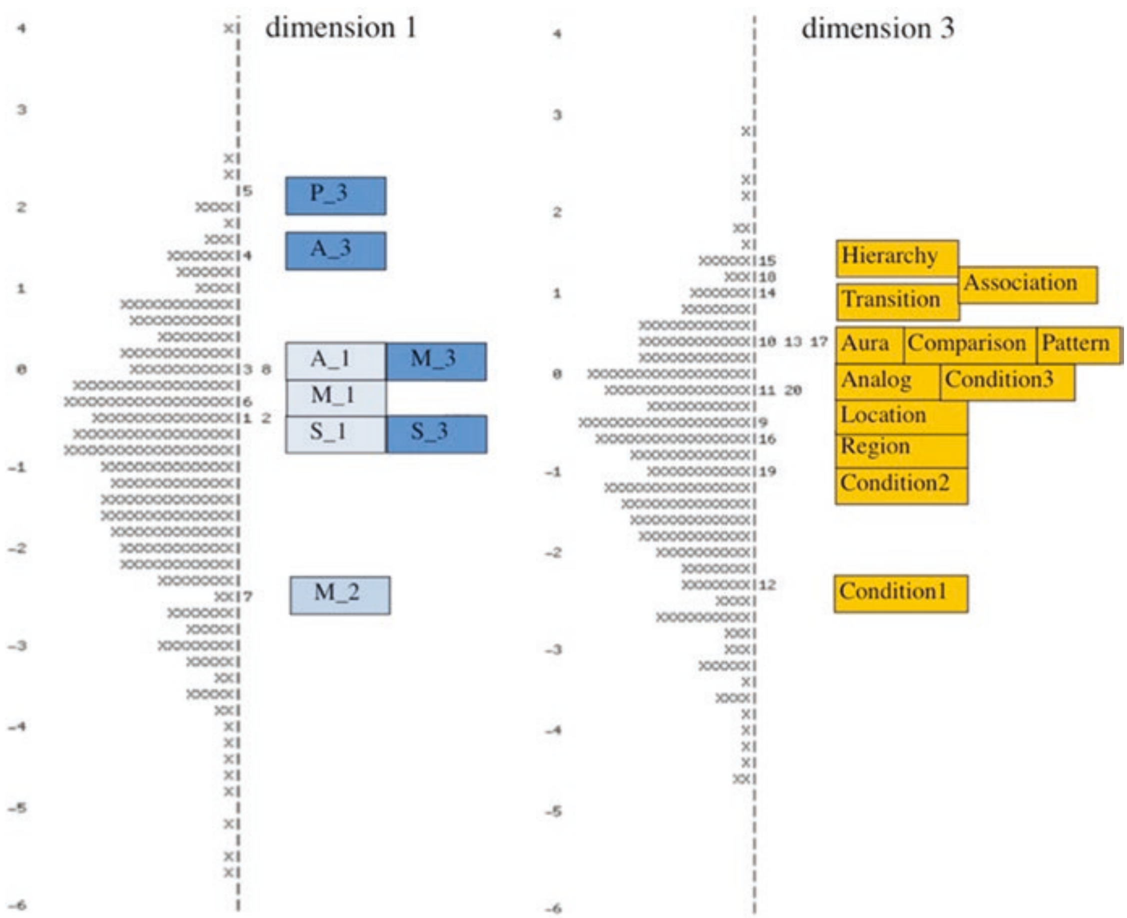

Fig. 4.4 Item map for Q1 without the excluded items
Dimension 1: the letter indicates the item stem, the number the assumed level
Items for "condition" in Dimension 3: a greater number indicates greater assumed complexity

according to complexity in the item map, as expected. These both point to future possibilities of a more qualitative graduation.

4.5 Second Quantitative Study (Q2)

4.5.1 Description of the Measurement Instruments

The second quantitative study used revised items. Q1 had shown that having to check several answers to get a credit was problematic. Consequently, in Q2, participants only had to choose one option to get full credit. Moreover, in contrast to earlier studies, the items were only drawn from one of the three areas of geography education specified by the national educational standards (DGfG 2010): that is, human-environment interaction. The items focused on the topic agriculture. Individual students, as well as a small number of experts were provided with draft versions of (some of) the items to get feedback and further improve the assessment instrument.

The questionnaire was implemented in Limesurvey. It contained background variables, a geographic pre-knowledge task asking students to write the names of seven countries marked on a world map, 13 items asking students to rate their own knowledge of different geographic aspects on a four-point scale, 13 items measuring their interest in these aspects on a five-point scale, and three items measuring interest in working with different media, on a five-point scale. This was followed by nine GSC tasks. Five tasks contained only Dimension 3 items, the other four tasks comprised both Dimensions 1–2 and Dimension 3 items. All in all, there were seven items for Dimension 1, five items for Dimension 2 and 11 items for Dimension 3. At the end, there was an open feedback field, as well as four statements (e.g., "The example countries were well chosen") that the students had to state their (dis)agreement to, on a five-point scale. A sample item can be seen in Fig. 4.5.

4.5.2 Dimensions of the Competency Model

Similarly to Q1, the three dimensions were first tested individually for one-dimensionality, with the help of a CFA in Mplus (Table 4.6) and a Rasch Analysis in Conquest. In the CFA, the Dimension 1 items fitted well to a one-dimensional model and also showed acceptable WMNSQ and *t*-values in the Rasch analysis. The Dimension 2 model did not converge in the CFA and thus had to be excluded from further analyses. For Dimension 3, the CFA showed that the 11 items had a bad model fit based on the CFI (comparative fit index), and an acceptable fit according to the RMSEA (root mean square error of approximation), but that the model fit

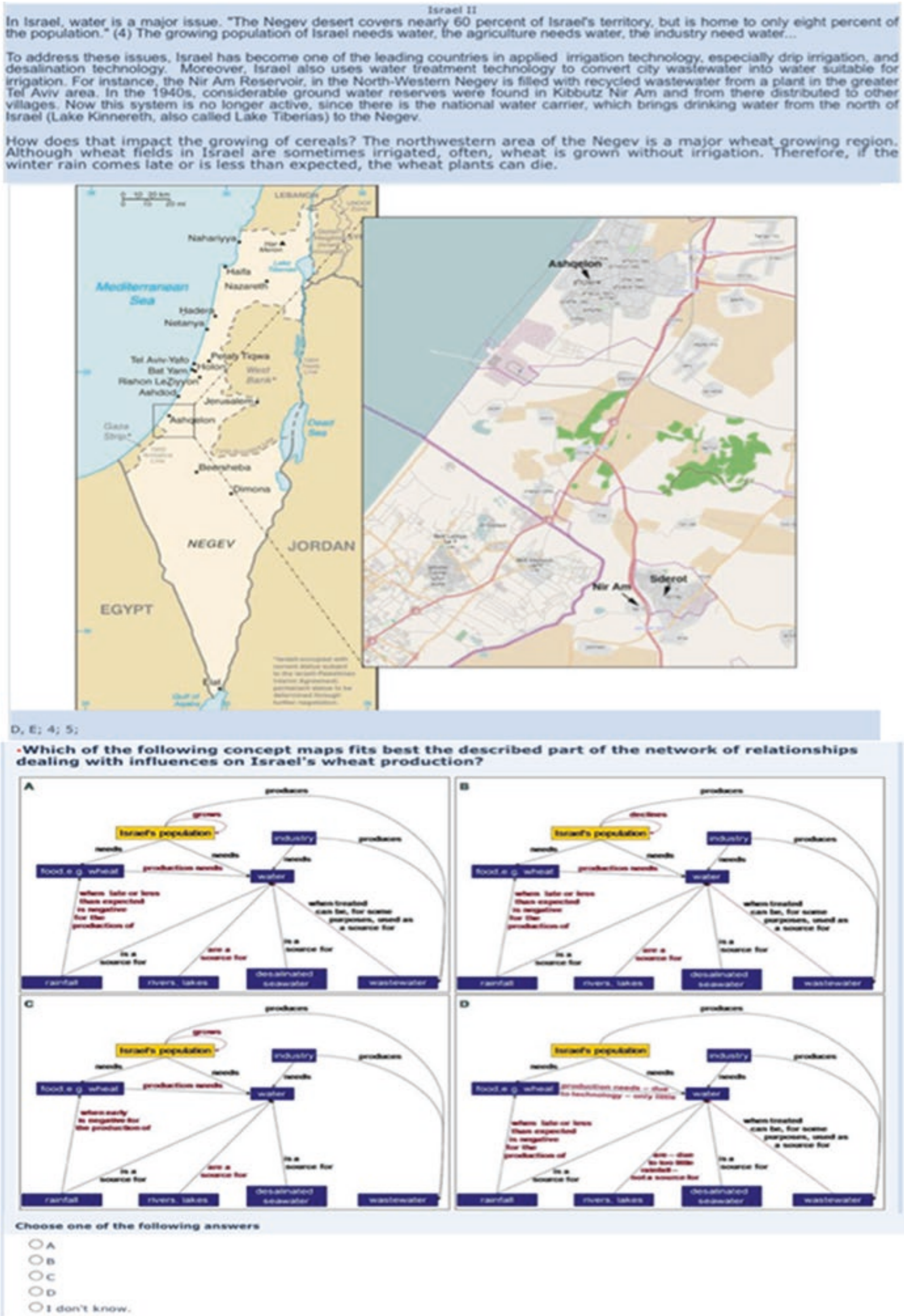

Fig. 4.5 Sample item from Q2: Item stem and one of the associated items (Dimension 1; English version)
Sources: images: (D) CIA World Factbook, (E) Open Street Map, text: (4) MFA: http://www.mfa.gov.il/MFA/InnovativeIsrael/Negev_high-tech_haven-Jan_2011.htm?DisplayMode=print

Table 4.6 Results of the separate tests for one-dimensionality for each dimension in Q2 (Mplus)

	Dimension 1: Comprehend and analyze systems	Dimension 2: Evaluate possibilities to act towards systems	Dimension 3: Spatial thinking
Number of items	7	5	11
Number of items without excluded items	7	–	6
χ^2	6.032		3.988
df	11		8
p	>0.10		>0.10
CFI	.99		.99
RMSEA	.00		.00
Conclusion	Items fit one-dimensional model well	Model did not converge; very low communalities ($h^2 > 0.02–0.07$)	Remaining items fit one-dimensional item well

could be greatly improved by excluding five items. The remaining items showed acceptable WMNSQ (weighted mean square) and *t*-values in the Rasch analysis.

Afterwards, a two- and a one-dimensional model were tested for the remaining Dimension 1 and 3 items. Both a one- and a two-dimensional model showed good fit values in the CFAs. The models did not differ significantly; thus, the one-dimensional model was preferred, due to parsimony. The Rasch Analysis in Conquest showed similar results (Table 4.7).

A possible reason for the differences from Q1 could be sample characteristics. In Q1, the sample had a slightly larger share of participants who had a very good GPA (grade point average) in the high school certificate (*Abitur*; Table 4.2). To test this hypothesis, the Q2 sample was split into a group with a GPA better than 2.5 (on a scale from 1 to 6, with 1 being the best and a 4 being considered a pass, $n = 173$) and a group with a GPA worse than 2.5 ($n = 151$).

The better than 2.5 GPA group did not show good model fit for both one- and two-dimensional models (Table 4.8). This seems to be caused by the items of Dimension 3 having very low communalities ($h^2 = 0.02–0.10$) and thus not constituting one factor. In contrast, the worse than 2.5 GPA group showed acceptable fit values for both models. The items of Dimension 3 constitute one factor ($h^2 = 0.10–0.76$). In the Rasch analysis, while for both groups the one-dimensional model was

Fig. 4.5 (continued) (5) http://www.ynetnews.com/articles/07340L-340392700.html, rest of the text based on: http://www.raymondcook.net/blog/index.php/2010/07/14/go-toisrael-drink-the-sea-israel-world-leader-on-desalination/, http://www.worldwatch.org/node/6544, http://tourguides0607.blogspot.com/2011/03/northern-negevtour.html, http://site.jnf.ca/projects/projectswater_reservoirs.html, http://www.jewishvirtuallibrary.org/jsource/judaica/ejud_0002_0015_0_14862.html, http://www.israelyoudidntknow.com/south-meansdesert/london-fires-negev-water/, http://site.jnf.ca/EDUCATIONSITE/jnf/negev3.html, http://mapsomething.com/demo/waterusage/usage.php, http://www.haaretz.com/news/low-rainfall-threatens-negev-wheat-and-golan-cattleranchers-1.207708

Table 4.7 Results of the tests for dimensionality (Dimensions 1 and 3) for the remaining items in Q2

Mplus	Two-dimensional model	One-dimensional model	χ^2 test of difference	
χ^2	21.644	21.392	χ^2	0.190
df	36	36	*df*	1
p	>0.10	>0.10	*p*	>0.10
CFI	.99	.99		
RMSEA	.00	.00		
Conclusion	One-dimensional to be preferred			
Conquest	Two-dimensional model	One-dimensional model	χ^2 test of difference	
Final deviance	3500.79	3502.75	χ^2	1.97
Number of estimated parameters	16	14	*df*	2
			p	0.374
Conclusion	One-dimensional to be preferred			

Table 4.8 Results of the tests for dimensionality (Dimensions 1 and 3) for the remaining items in Q2 by GPA higher (better) or lower (worse) than 2.5

Mplus	Two-dimensional model		One-dimensional model		χ^2 test of difference		
	High GPA	Low GPA	High GPA	Low GPA		High GPA	Low GPA
χ^2	35.360	18.810	39.065	18.475	χ^2	0.478	0.098
df	20	28	22	28	*df*	1	1
p	<0.05	>0.10	<0.05	>0.10	*p*	>0.10	>0.10
CFI	.38	.99	.31	.99			
RMSEA	.07	.00	.07	.00			
Conclusion	One-dimensional to be preferred						
Conquest	Two-dimensional model		One-dimensional model		χ^2 test of difference		
	High GPA	Low GPA	High GPA	Low GPA		High GPA	Low GPA
Final deviance	1763.79	1705.26	1766.40	1705.29	χ^2	2.61	0.03
Number of estimated parameters	16	16	14	14	*df*	2	2
					p	0.271	0.986
Conclusion	One-dimensional to be preferred						

to be preferred, there was a larger difference between the models for the better than 2.5 GPA group. Thus, while in both groups, the one-dimensional model had to be preferred, the results hint at some differences, especially with regard to Dimension 3. Therefore, the influence of GPA on competency structure should be further explored with a greater number of items for each spatial thinking skill.

Moreover, the sample also differed—to a much greater extent—with regard to the students' course of study (see Table 4.2). However, due to small cell sizes (e.g., for psychology students $n = 37$ in Q1 and $n = 58$ in Q2), separate models for psychology vs. geography education/geography students did not appear to be feasible.

4.5.3 Levels of the Competency Model

The GSC levels could only be examined on the basis of the remaining items of Dimensions 1 and 3. The item map, showing the difficulty of the items based on a one-dimensional Rasch model, can be seen in Fig. 4.6, which shows that the test was very easy for the sample.

For Dimension 1, similarly to Q1, levels could not be confirmed, because the Level 1 items were unexpectedly difficult. It is notable, however, that within the "N" item stem, dealing with New Zealand, the assumed levels were shown. Not every item stem had every level; thus, it cannot be confirmed whether there were systematic variations in difficulty between content areas or example countries.

For Dimension 3, an analysis of the raw data sum score of the remaining Level 3 items indicated the possibility of using the number of spatial thinking skills as a graduation. However, because of the exclusion of items, Level 3 could not be measured, with Level 1 being reached by 83.6 % of the respondents and Level 2 by 16.4 %. The item map (Fig. 4.6) shows differences in difficulty between the spatial thinking skills, pointing to future possibilities for a more qualitative graduation. However, a comparison with the item map from Q1 shows that item difficulty is not consistent, for instance, with hierarchy—the hardest spatial thinking item in Q1 and the easiest in Q2.

4.6 Discussion

The main aim of the studies was to explore the structure of GSC, especially with regard to the relationship between systemic and spatial thinking. The studies showed common results in part, but also differences.

4.6.1 Dimensions of GSC

Firstly, Dimension 2 could not be measured in the originally intended form of "acting towards systems", and had to be changed to "evaluating possibilities to act towards systems". Originally it had been planned to approximate the "acting" with MicroDYN items, which have proven useful in the assessment of interactive problem solving skills (see e.g., Greiff et al. 2013; Wüstenberg et al. 2012). However, the

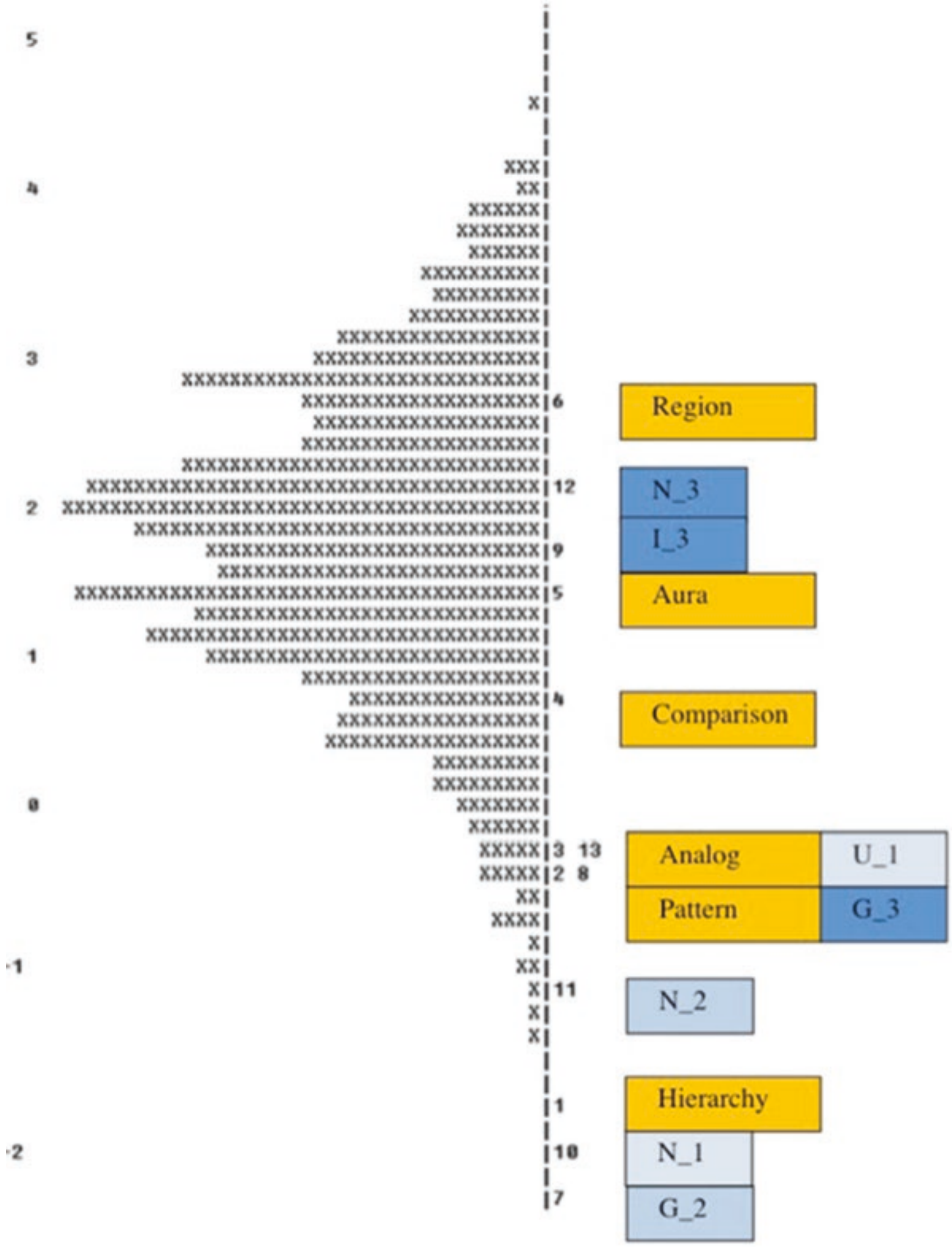

Fig. 4.6 Item map for Q2 without the excluded items
Dimension 1: the letter indicates the item stem, the number, the assumed level

CogLabs showed that MicroDYN items seem to be content-unspecific, making them not well suited to measuring geography-specific competencies. Even for systemic thinking as a general competency, however, there might be better-suited instruments than MicroDYN items, because systemic thinking means more than applying a VOTAT ("vary one thing at a time") strategy for exploration and drawing causal diagrams. For instance, the stock-flow failure reported by Sweeney and Sterman (2000) taps important issues in systems thinking that are not addressed by linear structural equation systems.

The revised Dimension 2 was not identifiable in both quantitative studies. This could be due to the small number of items or to the existing conception of this competency area. While learning to act adequately towards systems is a central part of geography education in Germany (e.g., DGfG 2010; Köck 1993), including competencies like being able to "[…] assess the natural and social spatial consequences of selected individual actions and think of alternatives" (DGfG 2007, p. 26), the studies showed some difficulties in measuring this dimension. Due to test time considerations, Dimension 2 should be removed from the competency model in further studies and rather treated separately, until a measurement scale for this dimension has been empirically validated, or the conception of that competency area is further developed.

Secondly, with regard to the other two dimensions, the studies showed differences. While Dimensions 1 and 3 fit a two-dimensional model in Q1, they fit both a two- and a one-dimensional model in Q2, leading to the preference of a one-dimensional model, due to parsimony. To further explore the reasons for these differences, several measures could be taken. Despite an item overhaul after Q1, several items for Dimension 3 had to be excluded in Q2. Further studies should employ a greater number of items for Dimension 3, including more than one item for each spatial thinking skill. This would lead to a longer test-time, however. Moreover, further analyses in Q2 hinted at a possible influence of GPA on competency structure, especially with regard to the fit of the Dimension 3 items. Consequently, a thinking-aloud study (CogLab) that comprised only spatial thinking items could be conducted with high and low GPA students, to investigate why for low GPA but not for high GPA students, the spatial thinking items constitute a homogenous factor. Additionally, the sample differed also with regard to other sample characteristics, such as the percentage of geography (higher in Q2), geography education (lower in Q2) and psychology students (lower in Q2). The items were designed to focus either on systemic thinking or on spatial thinking separately. However, in geographic inquiry, both are often interlinked, and thus might become more inseparable for students with a more extensive background in geography education. This could be further explored in expert-novice studies for instance. Furthermore, it might be helpful to have a third category of items that explicitly requires both spatial and systemic thinking skills.

Overall, the studies showed that in *geographic* contexts, systemic and spatial thinking are highly correlated or even inseparable. Thus, while studies focusing on just one of the two aspects *are* necessary for specific questions, they might only show a fragmented picture when it comes to understanding many geographic issues.

4.6.2 GSC Levels

Overall, the results of the levels are tentative till the structure of GSC is further explored. In general, Q1 was more difficult for the sample than was Q2, an effect possibly caused at least partly by the item formats used. Moreover, Q2 only used one broad topic area ("agriculture"), while in Q1, students had to switch between different topic areas.

In both quantitative studies, for Dimension 1, the remaining Level 2 items were consistently easier than the remaining Level 3 items. Level 1 items were shown to be more difficult than expected on the basis of the research literature (Ben-Zvi Assaraf and Orion 2005; Orion and Basis 2008). The reasons for this need to be explored. In general, there are recall items, for which students have to draw on their own pre-knowledge, or largely pre-knowledge-free tasks in which some or all information is given in the item stem. HEIGIS items belong to the second category. Thus, one avenue would be to compare both item types for the same topic and degree of complexity, to look at possible differences in level order. For recall tasks, it might be easiest to be able to name a concept as belonging to a sub-system (e.g., fertilizer has something to do with the topic "soil"), without remembering what exactly is the connection. In contrast, for item-stem-given information, to some degree participants might need to understand the information, before being able to decide whether a concept mentioned in the stem belongs to a sub-system or not. An alternative option would be to test one sample with both a translated version of the measurement instrument used by Ben-Zvi Assaraf and Orion (2005), and with the HEIGIS one. Another avenue could be, for instance, to have a substantial number of experts classify the items into the respective competency model components before the next study.

Additionally, new items and item formats could be tested and more control variables introduced, to explore potential effects of the kind of item, reading literacy or testwiseness. One promising item format seems to be multiple-select items, for which students have to check "right" or "wrong" for each choice (e.g., Pollmeier et al. 2011). This item format would ensure that students had to evaluate every single choice. It also could provide a closer link to pre-concepts, which is another option to improve the items.

For Dimension 3, quantitative graduation is one possibility. However, a more qualitative graduation would be preferable. In both studies, there is variation in difficulty between the items. However, as expected on the basis of Gersmehl and Gersmehl (2006), no general graduation can be observed. For instance, the spatial thinking skill "hierarchy" is very easy in Q2 but is among the most difficult spatial thinking skills in Q1. A greater number of items for each spatial thinking skill should be included, to find possible qualitative graduations and to test whether a one-dimensional scale comprising all spatial thinking skills is possible, as in both studies, some items had to be excluded. One possible graduation is complexity, as could be observed for the spatial thinking skill "condition" in Q1.

4.7 Conclusions

As outlined at the beginning, both systemic and spatial thinking are important aims of geographic education, but their relationship has not to any great extent yet been explicitly explored empirically. Hitherto, systemic and spatial thinking have often been studied separately (e.g., Battersby et al. 2006; Ben-Zvi Assaraf and Orion 2005; Lee 2005; Rempfler and Uphues 2012). The HEIGIS studies, however, show a close connection between systemic and spatial thinking when dealing with geographic systems. Consequently, while for some questions, focusing on either skill is necessary, both skills are needed in modeling GSC.

The studies also hint at some difficulties in measuring systemic and spatial thinking in geographic contexts. Thus, the model and associated items need to be further improved, to examine the relationship between both skills.

The HEIGIS studies were conducted predominantly with university students. The studies hinted at a potential influence of GPA on competency structure. A study by Orion and Basis (2008) showed an influence of grade on level order. In general, systemic thinking has been studied from kindergarten (see e.g., the project "Shaping the future" at the Heidelberg University of Education, http://www.rgeo.de/cms/p/pzukunft/) to postgraduate level. Ultimately, it would be helpful to have one model that can cover a person's whole learning history from beginner to expert, similar to what already exists in the area of foreign language learning (Council of Europe 2001). It should also comprise different interconnected versions, so that both a general competency overview and a more detailed picture for specific areas are possible within the same framework (see Viehrig 2015). This seems to be especially necessary in the light of current changes to the school system (e.g., the so-called *Gemeinschaftsschule* in Baden-Wurttemberg). Consequently, further studies should investigate the effect of grade level and GPA on both the structure and levels of geographic system competency to differentiate and improve the model. The HEIGIS studies showed that test time is a major constraining factor; this could be alleviated by using multi-matrix-designs for instance.

In summary, the project highlights the need for more research in this central area of geography education.

Acknowledgements This chapter is largely based on the final project report. The HEIGIS-project was funded by grants SI 877/6-1 and FU 173/13-1 from the German Research Foundation (DFG) in the Priority Program "Competence Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes" (SPP 1293). Thanks also to the student research assistants, who played an important role, especially with regard to the data collection in the CogLabs. Thanks also to all experts who provided valuable advice during item generation for all studies. Of course, thanks also to all participants.

References

Battersby, S. E., Golledge, R. G., & Marsh, M. J. (2006). Incidental learning of geospatial concepts across grade levels: Map overlay. *Journal of Geography, 105*, 139–146. doi:10.1080/00221340608978679.

Bell, T. (2004). Komplexe Systeme und Strukturprinzipien der Selbstregulation: Konstruktion grafischer Darstellungen, Transfer und systemisches Denken [Complex systems and structural principles of self-regulation: construction of graphical displays, transfer, and systemic thinking]. *Zeitschrift für Didaktik der Naturwissenschaften, 10*, 183–204.

Ben-Zvi Assaraf, O., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching, 42*, 518–560. doi:10.1002/tea.20061.

CoE (Council of Europe). (2001). Common European framework of reference for languages. http://www.coe.int. Accessed 22 Oct 2007.

DGfG (German Geographical Society). (2007). *Educational standards in geography for the intermediate school certificate*. Berlin: Author.

DGfG (German Geographical Society). (Ed.). (2010). *Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss—mit Aufgabenbeispielen* [Educational standards in geography for the intermediate school certificate—with sample tasks]. Bonn: Selbstverlag Deutsche Gesellschaft für Geographie.

Funke, J. (1990). Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen [System features as determinants of behavior in dynamic task environments]. *Sprache & Kognition, 9*, 143–154.

Funke, J., & Greiff, S. (2017). Dynamic problem solving: Multiple-item testing based on minimal complex systems. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 427–443). Berlin: Springer.

Gersmehl, P. J., & Gersmehl, C. A. (2006). Wanted: A concise list of neurologically defensible and assessable spatial thinking skills. *Research in Geographic Education, 8*, 5–38.

Gersmehl, P. J., & Gersmehl, C. A. (2007). Spatial thinking by young children: Neurologic evidence for early development and "educability". *Journal of Geography, 106*, 181–191. doi:10.1080/00221340701809108.

Greiff, S. (2010). *Individualdiagnostik der komplexen Problemlösefähigkeit* [Individual diagnostics of complex problem solving skills]. Münster: Waxmann.

Greiff, S., & Funke, J. (2009). Measuring complex problem solving: The MicroDYN approach. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (pp. 157–163). Luxembourg: Office for Official Publications of the European Communities.

Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in educational assessment: Analytical, interactive, and collaborative problem solving. *Journal of Problem Solving, 5*(2), 71–91. doi:10.7771/1932-6246.1153.

Hammann, M., Phan, T. T. H., Ehmer, M., & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education, 42*, 66–72. doi:10.1080/00219266.2008.9656113.

Kerski, J. J. (2013). Understanding our changing world through web-mapping based investigations. *J-Reading—Journal of Research and Didactics in Geography, 2*(2), 11–26. doi:10.4458/2379-02.

Klaus, D. (1985). Allgemeine Grundlagen des systemtheoretischen Ansatzes [General foundations of the systems theory approach]. *Geographie und Schule, 33*, 1–8.

Klieme, E., Hartig, J., & Wirth, J. (2005). Analytisches Problemlösen: Messansatz und Befunde zu Planungs- und Entscheidungsaufgaben [Analytical problem solving: measurement approach and results of planning and decision tasks]. In E. Klieme, D. Leutner, & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern* (pp. 37–54). Wiesbaden: VS.

Köck, H. (1993). Raumbezogene Schlüsselqualifikationen: Der fachimmanente Beitrag des Geographieunterrichts zum Lebensalltag des Einzelnen und Funktionieren der Gesellschaft [Space-related key qualifications: the subject-innate contribution of geographic education to the everyday life of individuals and functioning of society]. *Geographie und Schule, 84*, 14–22.

Lee, J. W. (2005). *Effect of GIS learning on spatial ability.* Doctoral dissertation, Texas A & M University, Texas. Retrieved from https://repository.tamu.edu/handle/1969.1/3896

Muthén, L. K., & Muthén, B. O. (2007). *MPlus user's guide*. Los Angeles: Author.

Orion, N., & Basis, T. (2008, March). *Characterization of high school students' system thinking skills in the context of earth systems*. Paper presented at the NARST annual conference, Baltimore.

Ossimitz, G. (1996). *Das Projekt "Entwicklung vernetzten Denkens": Erweiterter Endbericht* [The project "Development of networked thinking": Extended final report]. Klagenfurt: Universität Klagenfurt.

Ossimitz, G. (2000). *Entwicklung systemischen Denkens* [Development of systemic thinking]. Wien: Profil.

Pollmeier, J., Hardy, I., Koerber, S., & Möller, K. (2011). Lassen sich naturwissenschaftliche Lernstände im Grundschulalter mit schriftlichen Aufgaben valide erfassen [Can scientific achievements in primary school age be validly measured with written tasks]? *Zeitschrift für Pädagogik, 57*, 834–853.

Rempfler, A., & Uphues, R. (2010). Sozialökologisches Systemverständnis: Grundlage für die Modellierung von geographischer Systemkompetenz [Socio-ecological system understanding: foundation for the modeling of geographic system competence]. *Geographie und Ihre Didaktik, 38*, 205–217.

Rempfler, A., & Uphues, R. (2011). Systemkompetenz im Geographieunterricht: Die Entwicklung eines Kompetenzmodells [System competence in geographic education: The development of a competence model]. In C. Meyer, R. Henrÿ, & G. Stöber (Eds.), *Geographische Bildung. Kompetenzen in didaktischer Forschung und Schulpraxis* (pp. 36–48). Braunschweig: Westermann.

Rempfler, A., & Uphues, R. (2012). System competence in geography education. Development of competence models, diagnosing pupils' achievement. *European Journal of Geography, 3*(1), 6–22.

Rieß, W., & Mischo, C. (2008). Entwicklung und erste Validierung eines Fragebogens zur Erfassung des systemischen Denkens in nachhaltigkeitsrelevanten Kontexten [Development and first validation of a questionnaire to measure systemic thinking in sustainability-related contexts]. In I. Bormann & G. De Haan (Eds.), *Kompetenzen der Bildung für nachhaltige Entwicklung. Operationalisierung, Messung, Rahmenbedingungen, Befunde* (pp. 215–232). Wiesbaden: VS.

Smithson, P., Addison, K., & Atkinson, K. (2002). *Fundamentals of the physical environment*. London: Routledge.

Sommer, C. (2005). *Untersuchung der Systemkompetenz von Grundschülern im Bereich Biologie* [Examination of system competence of primary school students in the area of biology]. Doctoral dissertation. Retrieved from http://eldiss.uni-kiel.de/macau/receive/dissertation_diss_1652

Sweeney, L. B., & Sterman, J. D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review, 16*, 249–286. doi:10.1002/sdr.198.

Viehrig, K. (2015). *Exploring the effects of GIS use on students' achievement in geography*. Doctoral dissertation. Retrieved from http://opus.ph-heidelberg.de/frontdoor/index/index/docId/71

Viehrig, K., Greiff, S., Siegmund, A., & Funke, J. (2011). Geographische Kompetenzen fördern: Erfassung der Geographischen Systemkompetenz als Grundlage zur Bewertung der Kompetenzentwicklung [Fostering geographic competencies: Measurement of geographic system competence as foundation of evaluating the competence development]. In C. Meyer, R. Henrÿ, & G. Stöber (Eds.), *Geographische Bildung: Kompetenzen in didaktischer Forschung und Schulpraxis* (pp. 49–57). Braunschweig: Westermann.

Viehrig, K., Siegmund, A., Wüstenberg, S., Greiff, S., & Funke, J. (2012). Systemisches und räumliches Denken in der geographischen Bildung: Erste Ergebnisse zur Überprüfung eines Modells der Geographischen Systemkompetenz [Systemic and spatial thinking in geographic education: First results of testing a model of geographic system competence]. In A. Hüttermann, P. Kirchner, S. Schuler, & K. Drieling (Eds.), *Räumliche Orientierung: Räumliche Orientierung, Karten und Geoinformation im Unterricht* (pp. 95–102). Braunschweig: Westermann.

Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). *ACER ConQuest version 2.0. Generalised item response modelling software*. Camberwell: ACER Press.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence, 40*, 1–14. doi:10.1016/j.intell.2011.11.003.