DEVELOPMENT OF GEOMETRIC PROOF COMPETENCY FROM GRADE 7 TO 9: A LONGITUDINAL STUDY

Stefan Ufer
University of Munich, Germany
ufer@math.lmu.de

Aiso Heinze
University of Regensburg, Germany
aiso.heinze@mathematik.uni-regensburg.de

In this contribution we present an approach to describe the development of geometric proof competency from grade 7 to grade 9. Basis is a model on the complexity of geometric proof problems which integrates a mathematical and a cognitive perspective on proofs. Data collected in a three-year longitudinal study indicated that this model is appropriate to describe students' geometric proof competency in different grades in a cross-sectional manner as well as in its longitudinal development.

Type of paper: research paper Theme: cognitive aspect

1. Proof competency

In this contribution we investigate individual proof competency from a specific research perspective. Based on a quantitative empirical approach we try to describe the development of individual proof competency in relation to the complexity of proofs. In the center of our research is a specific conception of proof competency. According to Weinert (2001), competencies are defined as cognitive abilities and skills, which individuals have or which can be learned by them. These abilities and skills enable them to solve particular problems and encompass the motivational, volitional, and social readiness and capacity to utilize the solutions successfully and responsibly in variable situations.

Based on these assumptions the concept of proof certainly is important for the definition of proof competency. Our research interest is not to describe the students' understanding of mathematical proof explicitly but to describe their competency to solve proof problems. Therefore we take a normative perspective on mathematical proof. This view is influenced by the curriculum for secondary mathematics classroom and also includes criteria for the acceptance of students' proof as mathematically correct.

2. The complexity of mathematical proofs

In the last 20 years several quantitative empirical studies on geometric proofs competence of lower secondary students were conducted (e.g. Senk, 1985; Healy & Hoyles, 1998; Lin, 2000; Heinze, Reiss & Rudolph, 2005). Overall the results were disappointing because students had great difficulties to generate proofs even for simple geometric problems. Reasons for these moderate results in proof competency are various. Individual factors are, among others, lack of knowledge on geometric concepts, inadequate understanding of mathematical proof and difficulties in combining arguments to a logical chain. These findings are in line with results from mathematical problem solving research (cf. Schoenfeld, 1992).

Geometric proofs in the lower secondary level usually consist of one, two or three "proof steps". A proof step in this context is the application of a theorem known to the student. Many proof problems in lower secondary mathematics require the construction of only one proof step. Proofs with more than one step are usually not constructed stepwise, but ideas for all or most of the steps must be dealt with at the same time to build a plan for the proof. This encompasses the ability to understand that statements can have different status (for example, the hypothesis of the first step of a proof occurs as a premise for a second step) and to use this fact for creating deductive chains of arguments (cf. Duval, 1991). Thus proofs with more than one step require distinct additional cognitive processes and put qualitatively different demands on the student than proofs with only one step. In several studies it turns out that the number of “proof steps” required to prove a hypothesis is one central predictor for students’ performance on a geometric proof item (e.g., Reiss, Hellmich, & Reiss, 2002; Heinze, Reiss, & Rudolph, 2005).

Moreover, geometric proof problems can be distinguished from geometric calculation problems by the individual competency required for a solution. In this context geometric calculation problems are considered as problems in which the length of a side or the size of an angle must be calculated. As described in Heinze, Reiss, & Rudolph (2005) for grade 7 and 8 students there is a clear distinction in students’ performance on such geometric calculation items (competency level I), geometric one step proof items (competency level II) and proof items with more than one step (competency level III). These three competency levels correspond to achievement groups of the sample defined by the overall test score in such a manner that the low achieving students get acceptable results only for the calculation items, the middle achievement group for the calculation and the one step proof items and only the high achieving group of students is able to get acceptable results for multi step proof items.

3. Development of proof competency

When analyzing the development of (geometric) proof competency over three years (grade 7 to 9) two main aspects must be taken into account. We can assume, on the one hand, that during this time students acquire new (geometric) knowledge and, on the other hand, that they deepen the knowledge they already had before. From this follows that the students in grade 8 and 9 know more complex concepts and have more practice in procedures and heuristic strategies they learned in grade 7. Thus, we can expect that in general students in grade 8 and grade 9 perform better on an identical geometric proof test than they did in grade 7. At this point we want to state clearly that our considerations aim in investigating students’ competency on a system level. The quantitative methods as used in our studies are not appropriate to explain the competency of an individual student solving one specific proof problem. Longitudinal research by Küchemann and Hoyles (2006) gave evidence that there need not be a monotone increase of the performance in solving the same geometric proof problem over several years. However, investigating students’ competency at the system level of a school or a classroom and using an appropriate test consisting of several proof items which meet the requirements of probabilistic test models we can expect an increase of students’ performance. Nevertheless, we must take into account that we cannot describe specifically individual learning trajectories.

If we consider the aim of our study, a description of the development of geometric proof competency with the aid of a simple model of proof complexity then the question arises, whether the theoretical description of the complexity of proof items by proof steps is still sufficient. In our previous studies the model with three competency levels was confirmed in several cross-sectional studies in grade 7 and in grade 8 (overall sample encompasses more than 2000 students). However, in these studies different test were used in grade 7 and 8 for adapting the proof items to the curriculum. It is not clear how we can compare, for example, one step proof items based on content from grade 8 with two-step proof items based on content from grade 7 when investigating the proof competency of grade 8 students. It is even not clear, if a proof problem that requires from a mathematical point of view two arguments (two step item) also requires two steps when an individual processes this problem cognitively.

We assume that two different aspects of knowledge play a role when an individual solves proof problems. On the one hand, there are different qualities of conceptual knowledge which influence the fact whether the cognitive representation of a proof problem reflects all mathematical steps. On the other hand, strategic knowledge may play a different and independent role when planning a solution and combining arguments to a logical chain. Here we want to focus the cognitive representation of proof problems: In principal it is possible that an individual represents a two step item mentally as a one step item, because the first step is initial knowledge retrieved from memory in the sense of a chunk (cf. Anderson, 2004). In particular in geometry this can be the case, because frequently the initial situation of a proof problem (the prerequisites) is given in a graphical figure. Hence, possible geometric relations or facts are immediately cognitively present (e.g. the equal size of two vertical angles, the angle sum in a given triangle etc.) by recognizing specific aspects of a figure provided the individual has sufficient conceptual knowledge. If this is the case, then the individual does not need specific strategic knowledge to solve the two step problem, because she/he only needs one additional argument. An individual that is only able to represent the same item as a two step item probably needs strategic knowledge. As an example consider the item in figure 1. An immediate recognition of the vertical angle relation yields a mental representation of the proof item that requires only one additional argument about a straight angle. A student who just learnt the vertical angle relation may not be able to retrieve this conceptual knowledge from her/his memory and has to explore and to argue that the opposite angles are of equal size. Then the item is much more complex and strategic knowledge together with the ability to deal with different stati of statements (in the sense of Duval, 1991) is necessary.

4. A longitudinal study on geometric proof competency

To describe the development of geometric proof competency we conducted a study with students from grade 7, 8 and 9 (13-15 years old) in three consecutive years. The sample[1] comprises N = 1113 students for grade 7, N = 891 for grade 8 and N = 346 for grade 9. There is a core sample of N = 196 students that took part in all three measurements. All students visit the high attaining school track "Gymnasium" in Munich (Germany). The surveys were conducted using achievement tests with about 10 to 12 open-answer geometry items focusing on calculations, one step and multi step proofs (see section 4.1). The tests were administered in one mathematics lesson (45 minutes) by trained university assistants such that standardization was ensured.

4.1 Psychometric model and test instruments

In order to describe the development of proof competency from grade 7 to grade 9 it is necessary to have an underlying model which firstly describes the competency and secondly makes it possible to easily develop appropriate test items. For our study we started with our simple model on proof complexity and additionally take into account that the complexity of a proof item depends on the individual cognitive representation (see sections 2 and 3). This means that the allocation of an item on one of the three levels of competency may be different for different grades. However, this problem only occurs if one item is used in tests of different grades as an anchor item (see below).

The statistical approach in our study is to model the empirical data of all three tests uni-dimensionally by the dichotomous Rasch model (Rasch, 1960)[2]. The Rasch model is based on probabilistic test theory which allows the allocation of individual competency values and item difficulties on the same scale. One assumption in this model is that a person with competency

[1] The sample for grade 7 and grade 8 was part of a bigger research project. In grade 9 we conducted a smaller follow-up study.

[2] The scaling was performed with the software *ConQuest* (Wu, Adams & Wilson, 2006).

value x will solve an item with difficulty x by a probability of 50%. Working with probabilities on the one hand makes it possible to cover also specific individual behaviour as described for example by Küchemann and Hoyles (2006). On the other hand a probabilistic description makes it difficult to interpret the competency of a specific student. Here, all reported results are related to the system level.

For each grade we developed test items, such that all competency levels were covered by the test for this grade. The items for grade 7 and grade 8 were based on the corresponding curriculum in geometry. In grade 9 we only used test items related to the grade 7 and grade 8 curriculums. To link the three tests we used anchor items for every pair of tests. The anchor items of the study together with the core sample allow direct comparison of students' achievement at different times. As an example we will consider an anchor item that was administered in grade 7 and grade 9 (see figure 1). For grade 7 this item was assigned to competence level III because it consists of two steps. As a first step the angle between β and γ has to be identified as the vertical angle of α which has the same size. The second step uses the fact that angles at a straight line add up to 180°.

Prove $\alpha + \beta + \gamma = 180°$.

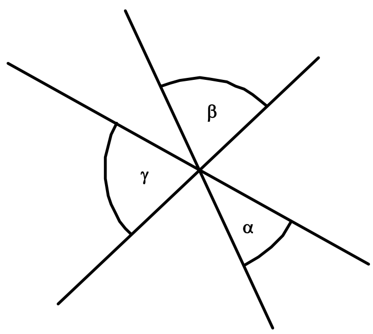

Fig. 1. Anchor Item Grade 7/9

A first analysis showed that about 10% of the students solved this item correctly in grade 7. This corresponds well to the assignment to competence level III. Nevertheless, in grade 9 almost a third of the students succeeded on this item. As described in section 3 we assume that for 9th graders the cognitive representation of the problem situation includes already the first step, because the vertical angle figure is recognized and retrieved as factual knowledge from memory. Based on this theoretical assumption, this item is transformed to a one step proof item of competency level II. Similar reassignments were made for other anchor items. Based on a reanalysis of the item contents we adjusted four of the nine anchor items to lower competence levels for the grade 8 or grade 9 measurements.

Item	**Level**	**Item Parameter**
10	1	*84.6*
06	1	*85.1*
02	1	*85.9*
03	1	*86.3*
01	1	*87.4*
15a	1	*88.5*
05a	2/1	*91.2*
08a	2	*93.3*
04	2	*93.6*
13a	2	*93.9*
11	1	*96.6*
09	2	*101.7*
17	2	*103.5*
15b	3/2	*103.9*
05b	3/2	*104.5*
07	3/2	*106.3*
08c	3	*108.6*
12	3	*110.3*
08b	3	*112.4*
15c	3	*113.0*
13b	3	*114.5*
16	3	*117.1*
14	3	*117.8*

Tab. 1. Item Parameters

Modelling the empirical data uni-dimensionally by the dichotomous Rasch model we obtained item and person parameters. The quality of the model fit was tested by checking item fit values for each item. For both procedures we got satisfying results, such that our test items are appropriate.

The item difficulties[3] span from 84.6 to 117.8, they corresponded to the assigned competence levels almost perfectly (see table 1). Those items assigned to different levels for different measurement points are located between the corresponding levels. Only item 11 turned out harder than expected for competence level I. This item is a calculation item, however, it requires an underlying argumentation with congruent triangles (see figure 2).

[3] The values of the latent trait are normed such that 100 is the mean of item parameters and the standard deviation of person parameters is 10.

Summarizing these results we can confirm that the model with three competency levels based on the item complexity is appropriate to develop test items for grades 7, 8 and 9. The proof complexity depends on the number of solution steps required by the proof problem in its cognitive representation. This means in particular that the question of proof complexity strongly depends on the individual (related to his/her knowledge). However, since there is a common basis of knowledge for students in the same grades (e.g. depending on the curriculum), in fact one can approximate possible cognitive representations of a proof problem and hence classify the complexity of proof items. At this point we want to mention that we benefit from the fact that our sample was recruited from German Gymnasiums, e.g. from the high attaining school track that is visited by about 40% of the students. We can assume that this sample is to some extent homogenous with respect to their mathematical knowledge such that an analysis of possible individual cognitive representations of proof problems is not too complicated.

The triangles ABC and A'B'C' are congruent with |AB| = |A'B'|. The points A, A', B and B' are on the same line. Determine δ. Show all your work!

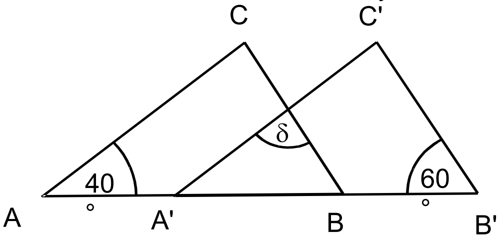

Fig. 2. Item 11 which requires argumentation and calculation.

4.2 Development of geometric proof competency

In this section we present findings about the development of geometric proof competency for the core sample (N = 196) from grade 7 to 9. The results presented here are based on the modelled empirical data.

4.2.1 Geometric proof competency from grade 7 to 9

First of all we compared the mean values of the individual test scores of each grade. Since all item parameters and all person parameters are allocated on one common scale, a direct comparison of the test scores is possible. The individual development of proof competency was tested by an ANOVA for repeated measurements. A significant effect was found for the time of measurement ($F(2,390) = 35.06$, $p < 0.001$, $\eta^2 = 0.154$). This indicates that the geometric proof competency significantly increased from grade 7 to grade 9. Absolute values and effect sizes are given in table 2. Since the overall mean value of the item difficulties was defined by 100, we can observe that the tests are somewhat to difficult for the sample (all mean values are below 100). However, for each grade we have no floor effect. The increase in proof competency for the first period between grades 7 and 8 is larger than for the second period. We assume that this effect is based on the instruction in geometric proof at the beginning of grade 8 and on the fact that the grade 9 test was based on the geometry contents from grades 7 and 8. Overall the increase between two measurement points (as effect size *d*) is comparable to the growth of mathematics achievement within one school year in Germany as it was shown in a study supplementary to PISA 2003 (e.g., Prenzel et al., 2006).

	M	SD	d	p
Grade 7	93.7	7.09	0.38	< 0.001
Grade 8	96.3	7.22		
Grade 9	98.3	7.67	0.27	< 0.001

Tab. 2. Development of proof competency

4.2.2 Geometric proof competency of different achievement groups

From an educational point of view it is interesting to consider the development of the competency for different achievement groups. Although we can observe a significant increase of geometric proof competence from grade 7 to 9 for the whole sample, the question occurs if all students benefit from this increase equally. It is a well-known fact that preknowledge is one of the strongest predictors for development of mathematical competency, such that one may expect that high achieving and low achieving students develop differently.

For this analysis we divided the core sample (N = 196) in three achievement groups due to their results in the grade 7 test. Then for each achievement group we computed the mean values for their competency in grade 7, 8 and 9. The results are displayed in the diagram in figure 3. It can be observed that the competency of the low achieving and middle achieving group develops very similar whereas the high achieving group is hardly able to improve its competency. This means that the increase from grade 7 to grade 8 for the whole sample goes back to the low and middle achieving students.

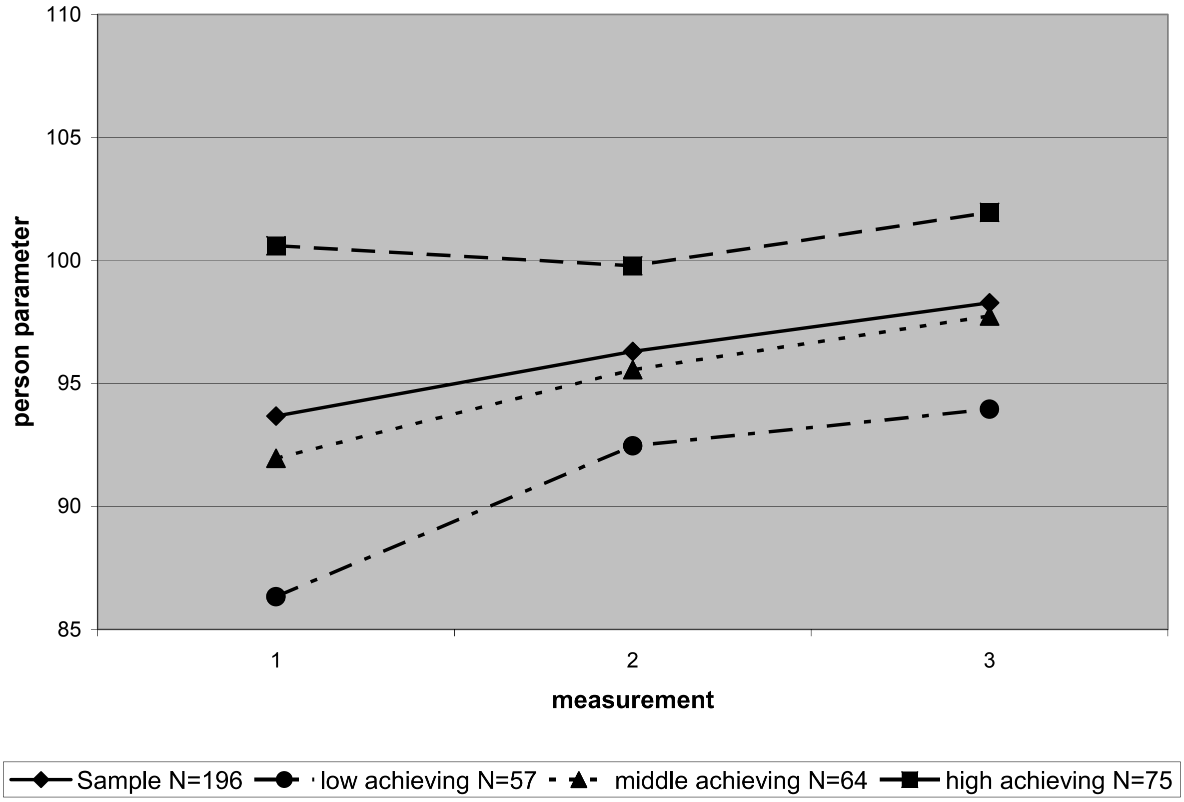

Fig. 3. Development of geometric proof competency for different achievement groups.

The fact that the high achieving students hardly show an increase in their competency cannot be explained by a ceiling effect. As we can observe in figure 3 even the mean values for the high achieving students are near the mean value of the item parameters which was defined as 100 for this scale[4]. Hence, the lack of improvement of this student group indicates that factors of mathematics instruction should be taken into account.

4.2.3 Geometric proof competency: qualitative improvement from grade 7 to 9

As presented in the previous sections the geometric proof competency of students increase from grade 7 to 9 though there are differences between different student groups. Based on our competency model on geometric proof competency we tried to analyze the quality of this improvement. We assume that students in grade 9, in general, can solve multi-steps proof items better than in grade 7. The open question is, if this effect goes back to the assumed fact that grade 9 students are able to represent two steps proof items cognitively as single step items such that they transform complex items to "simple" items or if can we in addition

[4] We want to mention again that the values are mean values. There are several students in the high achieving group which reach much better scores than 100.

conclude that 9th graders really improve their competency in multi-steps argumentation (in the sense of Duval, 1991, see section 2).

To analyze this question we consider separately for each grade the mean values for each competency level[5]. As we can observe in table 3 there is hardly an increase of students' ability in multi-steps argumentation from grade 7 to grade 9. Hence, we assume that the better performance of 9th graders is only based on their deeper conceptual knowledge which allows them to "simplify" proof problems by an adequate cognitive representation. The fact that there is no improvement may be associated to the result showed in figure 3 that the proof competency of high achieving students overall do not increase.

Percentages	competency level I	competency level II	competency level III
grade 7	76.6	55.8	23.4
grade 8	74.7	50.1	18.5
grade 9	78.7	42.9	15.8

Tab. 3. Mean values for different competency levels in different grades

5. Discussion

Summarizing the results we can state first that the students show a moderate performance in geometric proof problems. The mean values in table 2 show very clear that the person parameters are below the mean value of the item difficulties of the curriculum based test items. From a normative perspective these results are not satisfying for a sample of students from the high achieving school track. Overall our findings replicate results from several other studies cited before.

To describe the individual geometric proof competency we analyzed proof items with respect to their difficulty. Based on the work of Duval (1991) we assume that the number of proof steps an individual has to perform plays a central role. Hence, to consider the cognitive representation of a proof problem instead of the objective mathematical proof problem is the key point to approach the difficulty of proof items. Moreover, as described above, we postulate that there is no qualitative increase of the complexity from two-steps proofs to proof which requires more than two steps.

Based on these ideas we conducted the presented studies. Interpreting each measurement point as a single cross-sectional study the assignment of items to competency levels with respect to the item complexity (in its cognitive representation) reflects the item difficulty in each grade. Moreover, and this is one central result, the item complexity is still an important predictor when using test items comparing the proof competency in *different* grades. However, we have to adjust the assignment of items to competency levels with respect to the quality of geometric knowledge of the students. It turns out that the empirical item difficulty of these adjusted items is allocated just between two competency levels on the overall scale (cf. Table 1). We hypothesize that the individual quality of knowledge (in the sense of a more elaborated chunking) is one important factor for an increase of proof competency from grade 7 to grade 9. For the future we plan a specific study that investigates this relation between the quality of individual geometric knowledge and geometric proof competence.

Furthermore, there is no clear indicator that the individual ability in combining arguments or the individual strategic knowledge has significantly improved from grade 7 to grade 9 (see table 3). Comparing the performance for multi-steps proof items that fit the curriculum of the

[5] Here the anchor items are assigned to different competency levels depending on the grade. For example the item in figure 1 is assigned to level III for grade 7 and level II for grade 9. The item from figure 2 was assigned to level I for grades 8 and 9, since it is not a genuine proof item.

specific grades we cannot find an increase at all. It seems that the progress in geometric proof competency is mainly based on a better elaboration and chunking of the individual geometric knowledge. Combining this result with the finding that the high achieving group of students cannot improve their geometric proof competency from grade 7 to grade 9 (cf. figure 3) one can speculate that mathematics instruction did not provide adequate learning opportunities to these students. Since their geometric conceptual knowledge is already well elaborated these students may benefit most in learning strategies to explore multi-steps proof problems and to generate proof ideas which finally show them a way how to combine arguments to a proof.

References

Anderson, J. R. (2004). *Cognitive psychology and its implications* (6th ed.). New York: Palgrave Macmillan.

Duval R. (1991) Structure du raisonnement déductif et apprentissage de la démonstration. *Educational Studies in Mathematics,* 22(3), 233–263.

Healy, L. & Hoyles, C. (1998). *Justifying and proving in school mathematics*. Technical report on the nationwide survey. Institute of Education, University of London.

Heinze, A., Reiss, K., & Rudoph, F. (2005). Mathematics achievement and interest in mathematics from a differential perspective. *Zentralblatt für Didaktik der Mathematik, 37(3)*, 212–220.

Küchemann, D. & Hoyles, C. (2006). Influences on students' mathematical reasoning and patterns in its development: insights from a longitudinal study with particular reference to geometry. *International Journal of Science and Mathematics Education, 4* (4), 581 - 608.

Lin, F. L. (2000). An approach for developing well-tested, validated research of mathematics learning and teaching. In T. Nakahara & M. Koyama (Eds.), *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1 S. 84-88). Hiroshima: Hiroshima University.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J. & Schiefele, U. (Hrsg.). (2006). *PISA 2003. Untersuchungen zur Kompetenz-entwicklung im Verlauf eines Schuljahres.* Münster: Waxmann.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche.

Reiss, K., Hellmich, F., & Reiss, M. (2002). Reasoning and proof in geometry: prerequisites of knowledge acquisition in secondary school students. In A.D. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 113–120). Norwich: University of East Anglia.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). New York, NY: Simon & Schuster.

Senk, S. (1985). How well do students write geometry proofs? *Mathematics Teacher, 78* (6), 448–456.

Weinert, F. E. (2001). Concept of competence: a conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Seattle: Hogrefe & Huber Publishers.

Wu, M., Adams, R.J. & Wilson, M.R. (2006). *ConQuest. Generalized item response modeling software*. Melbourne: ACER.