R&E-SOURCE *http://journal.ph-noe.ac.at*
Open Online Journal for Research and Education
Special Issue #6, March 2017, ISSN: 2313-1640

An Alternative Approach for Measuring Gender Differences in Mathematical Sub-Competencies

TSG 52: Empirical methods and methodologies

Ann Cathrice George[*], *Alexander Robitzsch*[†]

Abstract

This paper presents a new perspective on measuring gender differences in the large-scale assessment study TIMSS. In contrast to existing approaches, the new one focusses on interactions between the students' mastery of mathematical sub-competencies. The new results may help to better understand the sources of differences and may thus lead to more targeted remedial actions.

Ein alternativer Ansatz zur Bestimmung von Geschlechterunterschieden in mathematischen Sub-Kompetenzen

Zusammenfassung

In diesem Artikel wird ein alternativer Ansatz zur Bestimmung von Geschlechterunterschieden in der Large-Scale Studie TIMSS vorgestellt und diskutiert. Im Gegensatz zu existierenden Methoden basiert die neue Berechnung auf Interaktionen zwischen dem Beherrschen einzelner mathematischer Sub-Kompetenzen. Die neuen Ergebnisse könnten dazu beitragen Geschlechterunterschiede in Mathematik besser zu verstehen und könnten daher eine empirische Grundlage bei der Entwicklung zielgerichteter Förderprogramme bilden.

Keywords:	*Schlüsselwörter:*
Mathematical sub-competencies	Mathematische Sub-Kompetenzen
Large-Scale Study TIMSS	Large-Scale Studie TIMSS
Alternative Measuring Approach	Alternative Berechnung
Cognitive Diagnosis Models	Kognitiv Diagnostische Modelle

1 Problem

The National Academy of Sciences (2006) reported that we still face underrepresentation of women in highly qualified occupational fields, especially techniques, science or engineering. A common denominator of these areas is mathematics, a subject which is already in schools known for its gender differences. In the German data set of the Trends in Mathematics and Science Study (TIMSS; Martin & Mullis, 2013) boys exceed girls by 12 points on the TIMSS scale in 2007 and by eight points in 2011. This is equivalent to a learning advantage of about a quarter of a learning year (Bos, Schwippert, & Stubbe, 2007), in TIMSS 2007 even more. Considering these differences with regard to the educational system and seeking to warrant equal starting positions for all students, a more detailed analysis is required. Such an analysis will deepen the understanding of potential sources of gender differences and consequently could lead to enhanced pedagogical courses of action.

---

[*] Institute for Educational Research, Innovation and Development of the Austrian School System (BIFIE); Salzburg, Austria.
Corresponding author. E-mail: a.george@bifie.at
[†] Leibniz Institute for Science and Mathematics Education (IPN), Centre for International Student Assessment; Kiel, Germany.

In that line, recent methodological developments allow for more differentiated approaches than reporting the differences on a general unidimensional modelled mathematical competence: For example, Brunner, Krauss and Martignon (2011) found the gender differences to vary across mathematical sub-competencies and thus suggest a separate reporting for each sub-competence.

1.1 Measuring gender differences in TIMSS

According to the TIMSS competence model for mathematics in the fourth grade, the students' general mathematical competence covers three content sub-competencies (*number*, *geometric shapes and measures*, and *data display*) and three cognitive sub-competencies (*knowing*, *applying*, and *reasoning*). Because students also require cognitive sub-competencies for solving content sub-competencies, the competence model combines both components. Following this principle, each TIMSS item covers exactly one content and one cognitive sub-competence. Hence, educational experts may assign each TIMSS item to exactly one of the nine cells in Figure 1.

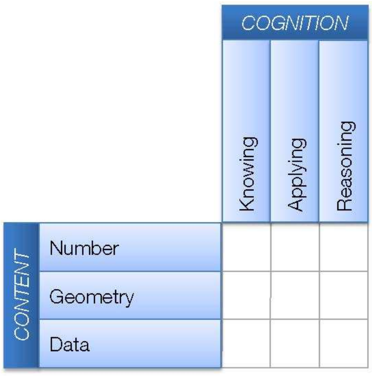

Fig. 1: Outline of the TIMSS competence model.

Based on the competence model, the official TIMSS reports do not only include gender differences for the general mathematical competence but also for each of the six sub-competences. The German 2011 cohort covers inhomogeneous gender differences in these sub-competencies: While boys score higher in the content sub-competencies *number* (+12 points) and *geometry* (+8 points), there was virtually no difference in *data* (−1 point). The gender differences in the cognitive sub-competencies vary between −4 points in *knowing*, 0 points in *applying* and 4 points in *reasoning*. These variations of the gender differences across different sub-competencies underline the necessity of a separate reporting.

The standard TIMSS procedure assesses gender differences in the content sub-competencies by applying a three dimensional model with each dimension representing one of the three content sub-competencies. Then for each content sub-competence a gender difference is determined (i.e. a marginal difference for each row in Figure 1). The identification of the gender differences on the cognition sub-competencies follows an analogous model approach and yields one gender difference for each cognition sub-competence (i.e. a marginal difference for each column in Figure 1). TIMSS officially reports only these six marginal gender differences (Mullis, Martin, Foy, & Arora, 2012). However, readers may as well construct the gender difference for each of the nine knot points between content and cognitive sub-competencies (i.e. for each cell in Figure 1) by simply averaging the respective marginal differences. In that line, Harks, Klieme, Hartig and Leiss (2014) found first evidence that combinations of sub-competencies reveal a more differentiated picture and thus have to be taken into account in order to fully understand gender differences in mathematics.

The TIMSS method of constructing gender differences for each field of Figure 1 assumes that a marginal gender difference equally affects the gender differences in the three knot points of the associated row or column. As a toy example, let us consider the following case, in which the numbers representing the gender differences may be interpreted as difference between the percentage of boys possessing the sub-competence and the percentage girls possessing the same sub-competence (cf. Table 1):

		COGNITION			
		Knowing	Applying	Reasoning	
CONTENT	Number	0.115	0.020	0.040	**0.030**
	Geometry	0.100	0.005	0.025	**0.000**
	Data	0.085	–0.010	0.010	**–0.030**
		0.200	**0.010**	**0.050**	

Table 1: Toy example for calculation of gender differences in TIMSS: The differences for each of the nine knot points between a content and a cognitive sub-competence are obtained by averaging the respective marginal gender differences (printed in bold). A marginal gender difference equally affects the gender differences in the three knot points of the associated row or column.

Regarding the cognitive domain, the gender difference in *knowing* is considerably larger than in *applying* and *reasoning* while in the content domain the gender differences are fairly equal for all three sub-competencies. The cells of Table 1 show the gender differences for the knot points calculated by averaging the respective marginal differences (printed in bold). The large marginal gender difference in *knowing* yields equally increased gender differences in the three knot points of the associated column. Analogous effects are obtained for the other two marginal cognitive and for the three marginal content sub-competences. A graphical illustration of these assumptions is characterized by parallel profile lines, see Figure 2, left hand side. Note, that the interpretation of the toy example stays valid if we think of gender differences in terms of differences between two values, one for boys and one for girls, on a unidimensional IRT scale.

However, the presented model assumes (in analysis of variance terminology) only main effects but no interaction effect to be present. However, this assumption is not explicitly tested – possibly it may rather be a concession to the goals of the TIMSS study, which is system monitoring but not diagnosing or explaining gender differences in great detail.

1.2 A new perspective on modeling gender differences

In order to reduce possible gender differences, we first have to understand their structure on a more differentiated level. Therefore, we drop the restriction that no interaction effect is present and allow the gender differences on the nine knot points to vary freely. For the toy example in Figure 2 this means, that the large marginal difference in knowing needs not necessarily affect all knot points *number/knowing*, *geometry/knowing* and *number/data* to the same extent. Rather, differences in *data/knowing* could be less distinct whereas the differences in *number/knowing* increase. The right hand side of Figure 2 illustrates one possible model exhibiting an interaction effect.

This article compares the two model variants, i.e. the TIMSS method and the suggested less restrictive approach. From a statistical point of view, we figure out, whether the more restrictive model suffices for describing the gender differences, or whether an interaction effect is required. Additionally, we discuss the substantial characteristics of the gender differences estimated in model approaches.

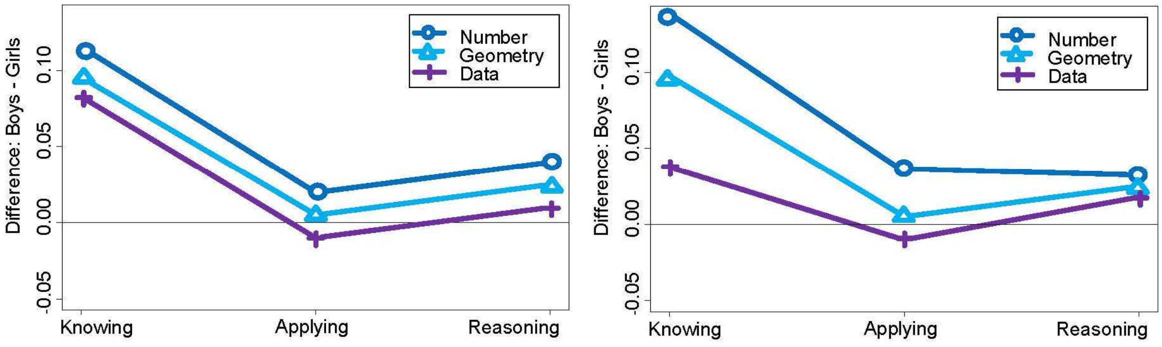

Fig. 2: Illustration of TIMSS method (left hand side) and new approach (right hand side) for calculation of gender differences in the nine knot points with data of toy example in Table 1. Left plot: The parallel profile lines are a result of the assumption that no interaction effects are present. Right plot: Interaction effects in a less restrictive model.

		COGNITION			
		Knowing	Applying	Reasoning	
CONTENT	Number	8	12	6	**26**
	Geometry	23	26	11	**60**
	Data	38	33	17	**88**
		69	**71**	**34**	**174**

Table 2: Assignment of analyzed 174 items to the nine knot points and six sub-competences.

2 Data and methods

The present study uses data from the mathematics part of the TIMSS 2011 study including 261339 fourth graders in 50 countries. Table 2 shows the overall count of the 174 analysed items in the nine knot points between content and cognitive sub-competencies (see TIMSS 2011 data material).

For estimating of the model variants, we employed a so-called cognitive diagnosis model (CDM; e.g. DiBello, Roussos, & Stout, 2007). Based on the individuals' responses to the items and the assignment of the items to a set of sub-competencies (in our case the nine knot points), the model determines, amongst others, the percentage of examinees possessing each of the knot points (for an introduction to CDMs see for example George and Robitzsch, 2015). For the present analysis, a specific CDM is required, the so-called Deterministic-Input Noisy "And" Gate model (DINA; Haertel, 1989). The chosen model supports the assumption that students should only be able to master an item if they possess both marginal sub-competencies (i.e. the respective content and cognition sub-competency) assigned to this item.

In a first step we estimated the less restrictive model by specifying a multiple group DINA model. In contrast to the officially reported TIMSS method, this model is less restrictive because it allows for interaction effects between the content and cognitive sub-competencies. In a second step, we conducted analysis of variance models (ANOVAs) to evaluate if the gender differences in the nine combinations of sub-competencies can (a) be already be explained to a sufficient extent by the content and cognitive sub-competencies or (b) need the modelling of an additional interaction effect. The resulting model of case (a) represents an analogue to the TIMSS approach, whereas case (b) supports our suggested model. Analyses were performed with the statistical programming framework R (R Core Team, 2017) using the R package CDM (George, Robitzsch, Kiefer, Groß, & Ünlü, 2016). Sample weights were taken into account in the estimation of the DINA model (George & Robitzsch, 2014). The ANOVAs were specified as Wald tests of the multiple group DINA model (cf. Johnson et al., 2013) in the CDM package by using the jackknife procedure for statistical inference.

3 Results

For a better understanding of the differences between the two model approaches, in this paper we restrict our analysis to two countries Germany and Austria. The German TIMSS sample includes 3995 fourth graders, while the Austrian data set contains 4668 students. The core statistic to be presented here is the difference between the percentages of boys and girls possessing each of the nine knot points. Positive values of this measure indicate advantages for boys.

3.1 Germany

For the German data set, Figure 3 shows the differences between the model analogue of the TIMSS approach (left hand side) and our suggested less restrictive approach (right hand side). The gender differences reported in the TIMSS analogue (a) are comparable to the official TIMSS results (Mullis, et al., 2012; Brehl, Wendt, & Bos, 2012): Boys have advantages in *number* and *geometry* while girls have advantages in *data*. As discussed before, the gender differences in the cognitive sub-competencies have equal effects on all content sub-competencies resulting in parallel profile lines.

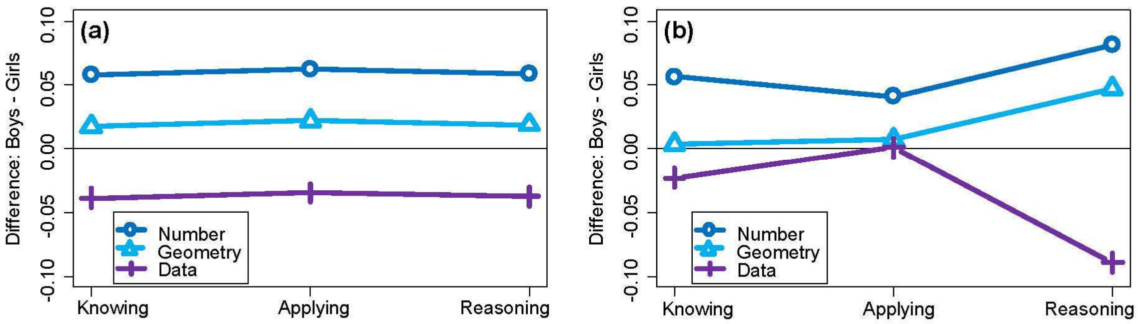

Fig. 3: Gender differences for German data modeled (a) following an approach analogous to the TIMSS variant (left hand side) and (b) less restrictive including interactions between content and cognitive sub-competencies (right hand side).

The results of our less restrictive approach (b) – the one modelled by the DINA model – differ significantly ($p < .05$) compared to the standard approach (a). Still, boys show advantages in *number* and *geometry* and girls in *data*. Also, the gender differences in knowing are comparable to the ones measured before. However, while the order of the differences remains constant in *applying* we find a marked shrinkage of differences. On the other hand, again with constant order, we observe an increase of differences in *reasoning* compared to *knowing*.

3.2 Austria

Figure 4 shows the differences between the model approaches (a) and (b) for the Austrian data. Again the gender differences in the TIMSS analog (a) are comparable to those of the official report (Mullis et al., 2012): Boys have advantages in all three content sub-competencies which are most prominent in *data*. Again the results of the DINA model (b) differ significantly ($p < .05$) compared to (a). In line with (a) model (b) shows the smallest advantages for boys in the content sub-competence *knowing*. In contrast to (a) the advantages for boys in *geometry*/*knowing* increase while they decrease in *geometry*/*applying*. The gender difference in *reasoning*/*data* changes most, as approach (b) now certifies girls advantages.

3.3 International

In both countries the gender differences obtained by the standard approach (a) and the new model approach (b) differ significantly (In the whole study holds significant differences between the two model approaches in 40 of the 50 participating countries).

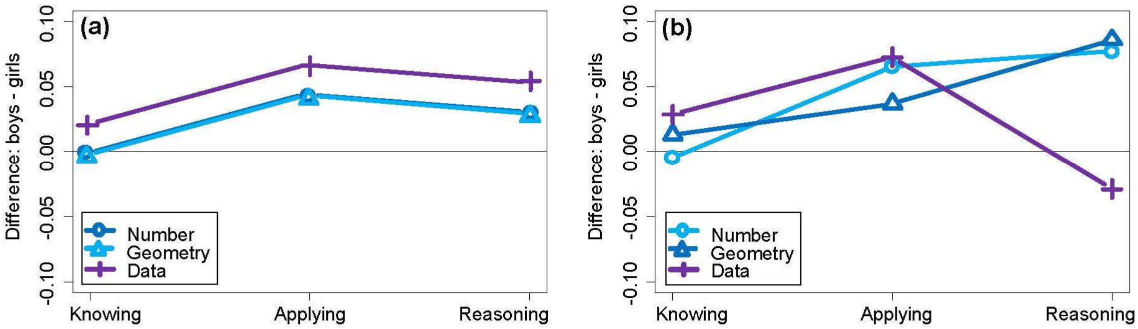

Fig. 4: Gender differences for German data modeled (a) following an approach analogous to the TIMSS variant (left hand side) and (left hand side) and (b) less restrictive including interactions between content and cognitive sub-competencies (right hand side).

However, the type of change between the two models differs in the two countries considered here. In Germany, the order of the gender differences in the content sub-competences remains the same in both model approaches, while the cognitive dimensions induce a shrinkage or expansion. In contrast, the order of the gender differences in content domains changes in Austria. In both countries the new model approach assigns girls advantages in *reasoning*/*data*, which are less striking in the standard approach (a).

4 Discussion

Since the size of the gender differences varies between the standard TIMSS procedure (the model (a) analogue) and our alternative model approach (b), the estimation strongly relies on a theoretical assumption: Are we assuming that (a) effects in one content sub-competence similarly affect all cognitive sub-competence and vice versa, or do we suppose (b) effects in the content sub-competences to individually influence each cognitive sub-competence? The selection of the model approach may not only be decided based on statistical measures as significance and model fit, but also based on the goal of the study (Pellegrino, 2001). If the primary goal is system monitoring, the detailedness of the results in model (a) may suffice; but if one is interested in explaining the reasons of gender differences, model (b) may provide a more accurate empirical basis. Substantially, our more fine-grained alternative model (b) directly reflects the concepts of mathematical competence models and their cognitive psychological foundation (Guilford, 1967; NCTM, 2000): These theories give evidence for mathematical abilities being a cluster of a content dimension, a process dimension, and a difficulty dimension (cf. also Roppelt, Blum, & Pöhlmann, 2013). Hence, model (b) may better support linking empirical results to substantial explanations.

Note that modelling the interaction of content and cognition dimension could also be conducted by assessing multidimensional item response models (MIRT) with continuous competencies (as opposed to dichotomous competencies in the CDM approach). In such a nine-dimensional MIRT approach, each dimension represents one of the nine knot-points. Gender differences could be specified for each dimension. Alternatively, we could specify a unidimensional model for an overall mathematics competency in which the gender differences in the nine knot-points are assessed by investigating a differential item bundle functioning approach. In this approach the nine knot-points are defined as item bundles (see for example Douglas, Roussos, & Stout, 1996).

In either case, the results of gender differences in model (b) generate entirely new questions for educational experts of mathematics. For example, in Germany, we could further explore why gender differences in applying are smaller compared to those in *reasoning*. Or, in both countries, one could try to explain the advantages of girls in *data*/*reasoning*. In doing so, one could analyse item contents in connection to students' responses in order to generate hypothesis about possible reasons for the differences. Following this direction should allow a targeted development of remedial actions.

The alternative model approach presented here allows for more general applications than those presented here: The comparison of boys and girls could be extended to various groups of interest, like, for example, migrants and non-migrants or even between groups of migrants with different cultural backgrounds. Furthermore, our modelling approach also allows even more detailed results. For example, the analysis of the combinations in which students are likely to possess the various sub-competencies (cf. Johnson et al., 2013), offers an empirical base for eliciting assumptions about possible ways in that students acquire knowledge in mathematics.

References

Bos, W., Schwippert, K., & Stubbe, T. C. (2007). Die Kopplung von sozialer Herkunft und Schülerleistung im internationalen Vergleich [Linking social background and student abilities: An international comparison]. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, & R. Valtin (Eds.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 225–247) [*IGLU 2006. Reading abilities of fourth graders in Germany and in an international comparison*]. Münster: Waxmann.

Brehl, T., Wendt, H., & Bos, W. (2012). Geschlechtsspezifische Unterschiede in mathematischen und naturwissenschaftlichen Kompetenzen [Gender specific differences in abilities of mathematics and science]. In W. Bos, H. Wendt, O. Köller, & C. Selter (Eds.), *Mathematische und naturwissenschaftliche*

Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich (pp. 203–230) [*Students abilities in mathematics and science in Germany and in an international comparison*]. Münster: Waxmann.

Brunner, M., Krauss, S., & Martignon, L. (2011). Eine alternative Modellierung von Geschlechtsunterschieden in Mathematik [An alternative model approach for gender differences in mathematics]. *Journal für Mathematik-Didaktik*, 32, 179–204.

DiBello, L., Roussos, L., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of Statistics*, Volume 26, Psychometics (pp. 979–1030). Amsterdam: Elsevier.

Douglas, J. A., Roussos, L. A., & Stout, W. (1996). Item-bundle DIF hypothesis testing: Identifying suspect bundles and assessing their differential functioning. *Journal of Educational Measurement*, 33, 465–484.

George, A. C. & Robitzsch, A. (2014). Multiple group cognitive diagnosis models, with an emphasis on differential item functioning. *Psychological Test and Assessment Modeling*, *56*(4), 405–443.

George, A. C. & Robitzsch, A. (2015). Cognitive diagnosis models in R: A didactic. *The Quantitative Methods for Psychology*, *11*(3), 189–205.

George, A. C., Robitzsch, A., Kiefer, T., Groß J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis modeling. *Journal of Statistical Software*, *74*(2), 1–24.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26, 301–323.

Harks, B., Klieme, E., Hartig, J., & Leiss, D. (2014). Separating cognitive and content domains in mathematical competence. *Educational Assessment*, 19, 243–266.

Johnson, M., Lee, Y.-S., Sachdeva, R. J., Zhang, J., Waldman, M., & Park, J. Y. (2013, March). *Examination of gender differences using the multiple groups DINA model*. Paper presented at the 2013 Annual Meeting of the National Council on Measurement in Education, San Francisco CA.

Martin, M. O. & Mullis, I. V. (2013). *Methods and procedures in TIMSS and PIRLS 2011*. Lynch School of Education, Boston College: TIMSS & PIRLS International Study Center.

Mullis, I., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international report*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Academy of Sciences (2006). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. The National Academies: Washington, DC.

NCTM, National Council of Teachers in Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know*. National Academic Press: Washington DC.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

Roppelt, A., Blum, W., & Pöhlmann, C. (2013). Die im Ländervergleich 2012 untersuchten mathematischen und naturwissenschaftlichen Kompetenzen [Abilities in mathematics and science established in the comparison of federal states]. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle, & C. Pöhlmann (Eds.), *IQB Ländervergleich 2012: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (pp. 23–37) [*IQB comparison of federal states: Competences in mathematics and science at the end of the tenth grade*]. Münster: Waxmann.