Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Wright-Map to investigate the actual abilities on math test of elementary students

To cite this article: H Hilaliyah *et al* 2019 *J. Phys.: Conf. Ser.* **1318** 012067

Wright-Map to investigate the actual abilities on math test of elementary students

H Hilaliyah[1], Y Agustin[1], S Setiawati[1], S N Hapsari[1], I B Rangka[1,*] and M Ratodi[2]

[1]Universitas Indraprasta PGRI, Jl. Nangka No. 58 C (TB. Simatupang), Kel. Tanjung Barat, Kec. Jagakarsa, Jakarta Selatan 12530, Jakarta, Indonesia
[2]Universitas Islam Negeri Sunan Ampel, Jalan A Yani No. 117, Jemur Wonosari, Wonocolo, Surabaya, Indonesia

*itsarbolo.rangka@unindra.ac.id

Abstract. The evaluation of learning outcomes, especially class promotion tests, is high possibility vulnerable to biases that cause students ability to be described inappropriately. Our study proposes to investigate students abilities based on math test in elementary school. This study involved 31 elementary school students. We used Wright-Map on Rasch modelling as an objective way to measuring students actual abilities. Data were collected based on mathematical tests on basic arithmetic, fractional and geometry. Based on Wright-Map found 11 students who have above average mathematical ability. In contrast, 13 other students had below average math skills. Unfortunately, there are 7 students who cannot be measured mathematical ability because of Outfit exposed. The other facts, overall students have difficulties in fractional matter in comparative and scale problems, but students have no difficulty in operations to change fractions to form percent and decimal. The comparison of mathematical ability between male and female students is also an interesting issue that we reviewed in this article. By using Wright-Map the student's actual ability can be drawn up accurately. It is very powerful for classroom teachers to learn student's qualification in classroom promotion and/or student grade ratings in mathematics materials.

1. Introduction

The development Physics as "the queen of sciences" has brought issues to the measurement model of social science. How can a personal experience have a unit of measurement common to different persons? Unfortunately, the question is "problematic" to answered especially to social scientists until the evolution of Rasch (Family of Item Responses Theory) modeling is expected to contribute quality measurements as provided by natural sciences [1-5]. The measurement model in education is dominated by Classical Test Theory (also known as true score theory based on raw score). The basic assumption of Classical Test Theory is that the score *(X)* occurs from a true score *(T)* and error measurement *(E)*. This means, in the score obtained by a student there is a true score and error measurement. The visible score *(X)* is observed, but the true score *(T)* and error measurement *(E)* are Latent. Furthermore, using Classical Test Theory to determine students' learning achievement has weaknesses, specifically: (1) raw scores are not essentially measurements; (2) the raw score is just the initial information; (3) the raw scores have weak quantitative meanings; (4) the raw score did not indicate a person's ability to the task being performed, and; (5) raw scores and percentages of correct answers are not always linear [6-9]. The Rasch Model differs from true score theory. First, it is expressed at the item level and the instrument level, not just the instrument level, as is true score theory. Second, it concentrates awareness on modeling the probability of the observed responses, rather than on modeling the responses [9]. Rasch Model implementation also reaches analysis for test formats presented in the form of Dichotomy, Polytomy, and Partial Credit Model (PCM) [10]. Furthermore, another impressive concept in Rasch modeling is the practice of Wright-Map [11,12] to illustrate the attribute constructs being measured. Wright-Map conforms to the phrase "a picture is worth more than thousand words" [13], making it clearer for readers to find out the quality of measurements was performed.

Wright-Map is a visual representation to recognize the existence of items and respondents in one dimension of measurement [14]. This means Wright-Map provides a comprehensive and meaningful context on the measured construct and content that consists of the quality respondent (lowest to highest) and the quality of item (easiest to the most complex) [6,15]. Wright-Map provides the idea that the higher of cognitive processes must be correlate with the higher of the difficulty level of the items that measure. As an example, referring to Bloom's taxonomy students who are skilled to handle on synthetic exam items can surely have greater capabilities than students who are only able to achieve on items that are memorable [16,17]. Studies show the Wright-Map application can help show the location of items and people; as an example of research conducted by Knoch which uses Wright-Map to test the effectiveness of feedback on the rating behavior [18]. Study conducted by Wilson indicates that Wright-Map can be used in cognitive diagnosis [19]. Wright-Map allows classroom teachers to assess how effectively the test items are describing a variable or to analyze the predicted order of item difficulty with the actual order of item difficulty [15]. In the context of achievement testing, we would say that the ability of the respondent is (a) equal to, (b) greater than, or (c) less than the difficulty of the item [9]. The purpose of Wright-Map not only illustrates the construct into a picture; Moreover, Wright-Map is also put as an empirical fact that emerges from an objective measurement. The left panel on the Wright-Map location of the respondent based on his ability, and the right panel is the location of the item based on the difficulty level. On the other hand, the practice of Wright-Map provides an awareness of classroom teachers to reconstruct the quality of instruments that have been developed and/or to create improvements where necessary [6]. Furthermore, generating Wright-Map also means we can get precise information linked to (1) the value and location of logit persons and items, (2) the ability of items to measure student's ability, or vice versa, and (3) identify students' abilities based on certain demographics (gender, place of residence, or other characteristics).

The first aim of this article is to evaluate the quality of mathematics tests and student consistency during the exam. Second, set up the Wright-Map to figure the exact location of an item based on its complexity and the location of the person based on the actual ability it has. Third, evaluate the students' mathematical abilities based on gender. This article gives information revealed to the practice of learning assessments that are adequate to demonstrate different points of view in determining the characteristics and abilities of students in carrying out the exam.

2. Method

The study linked to 31 students in elementary school at Jakarta. Students who are the subject of this study are students who received the math test. Learning materials of math tested include (1) basic arithmetic, (2) fractional, and (3) geometry. The three materials were developed into 25 items Multiple Choice Questions (MCQ). Data analysis was conducted using Rasch Modeling [3,10,20] and Wright-Map generated by WINSTEPS 4.1.0 as Rasch Measurement Computer Programs [21].

Data analysis phases established by conducting (1) initial screening; (2) generating a Wright-Map, and; (3) interpreting mathematical ability comparisons between males and females. The results of data analysis as whole can be accessed through https://osf.io/mabzw/ Open Science Framework [22]. All information regarding the students as respondents in this research is credential and is managed responsibly. Furthermore, there is no benefit provided by researchers and/or schools for student involvement in this study.

3. Results and discussions

3.1. Initial screening: class promotion test on math field

The high potential of biases against the mathematical test instrument provided in the classroom is a major concern of us. Student's actual ability particularly be identified if the instrument applied has superior psychometric properties, and for that we organize an initial study of the quality of the mathematics test instrument addressed to the students.

The reliability aspect proves that the mathematics test instrument has a remarkably strong reliability (α =.90). Unfortunately, this is not paired by the consistency aspect of the students reach during the exams (α = .64). This is also supported by the interaction value between the test items and the students who are in moderate categories (α = .67, SEM = 1.90). Furthermore, the realization of item information unit in math test only reached 41.6%. Nevertheless, this is considered acceptable because the percentage of it exceeds 40%, which means that the mathematical test applied is unidimensional [3].

Table 1. Item and person outfit (N=31, I=25).

Items #	Outfit MNSQ	Person	Outfit MNSQ
Item 18	1.98	12L	2.77
Item 24	1.55	14P	1.60
Item 21	.41	29P	1.53
Item 04	.30	02P	.48
		21P	.40
		30P	.39
		08L	.31

L = *Laki-laki* (Male), and P = *Perempuan* (Female). Measure in Logits

Furthermore, if we seek closely at which items and students are not completing well, particular items and person out of substantial function or Outlier-Sensitive Fit (outfit) against Mean-Square Fit Statistic (MNSQ).

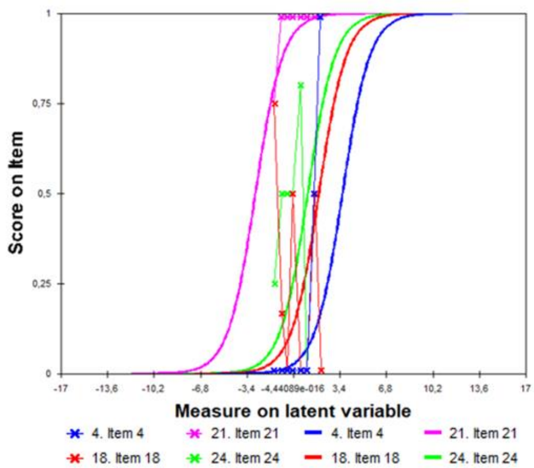


Figure 1. Item Characteristic Curves (ICC) of 4 outfit items.

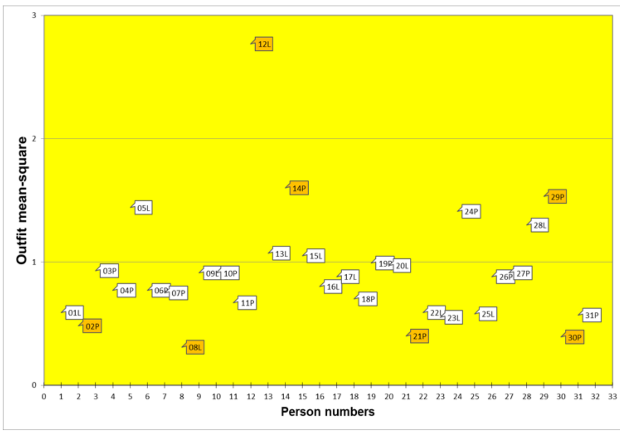


Figure 2. Head map of 7 outfit students.

Items and individuals are classified outfit if they have MNSQ outfit out of the strongest condition of measurement (+0.5 till +1.5 logit). Based on Table 1 we recognized that there are 4 items, and 7 students are outliers. The four items are not suited for measuring students' math skills. On the other hand, 7 students indicated a pattern of response that didn’t match with their actual capabilities. Visualization of outfit items and person cases is illustrated in Figures 1 and Figure 2. The competence materials measured in misfit items are (1) Using fractions in comparison and scale problems (Item 04 and Items 18); (2) Sums and subtracts various fractional forms (Item 21), and; (3) Identify flat wake properties (Item 24).

Based on the findings at the initial screening stage, we decided to eliminate misfit items and persons to be inspected using Wright-Map; or in other words entirely 21 Items and 24 students who measured their ability in a mathematical test that has been implemented.

3.2. Wright-map: students actual ability vs item of math test

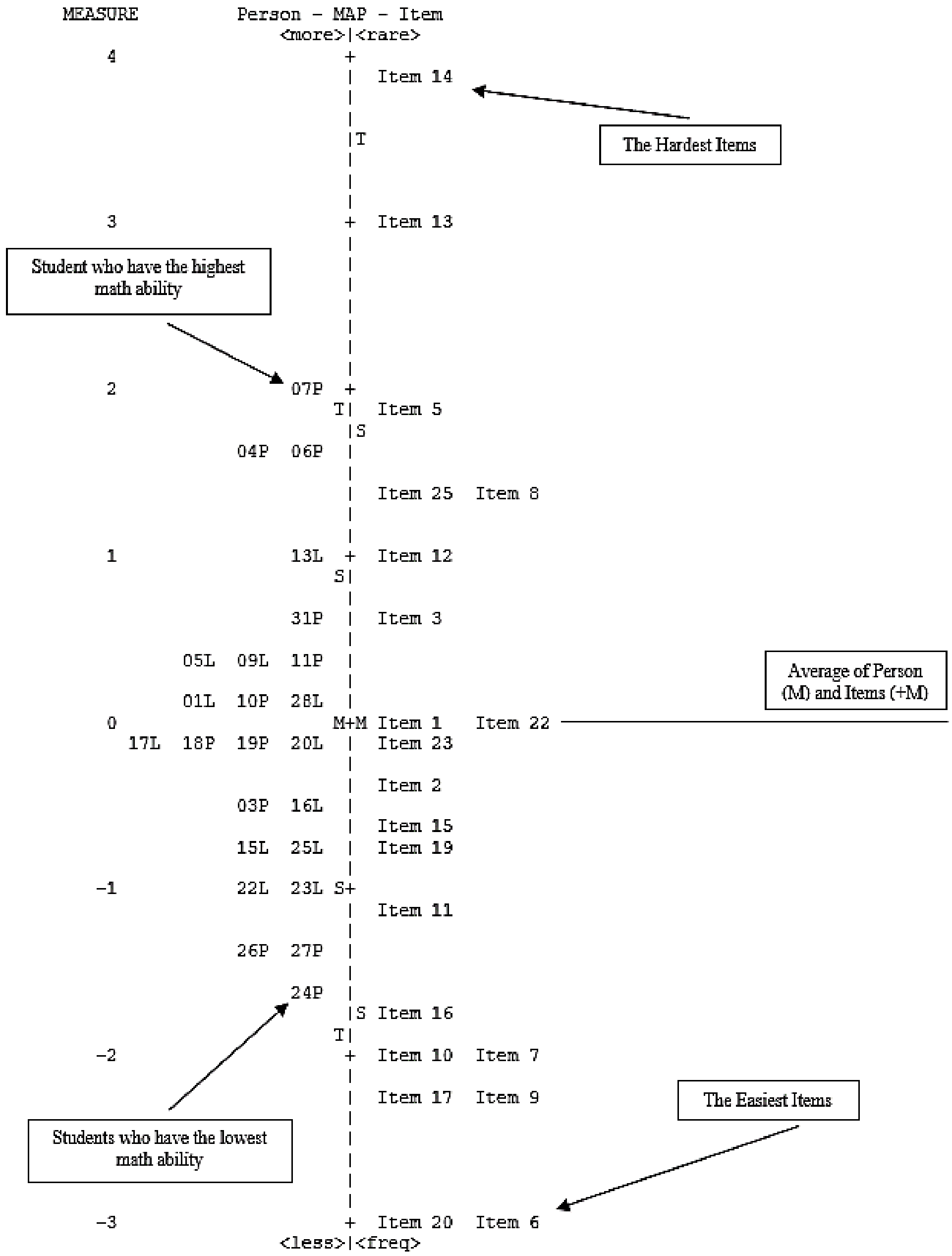


Figure 3. Wright-map of items and persons on math test; rasch measurement (N=24, I=21).

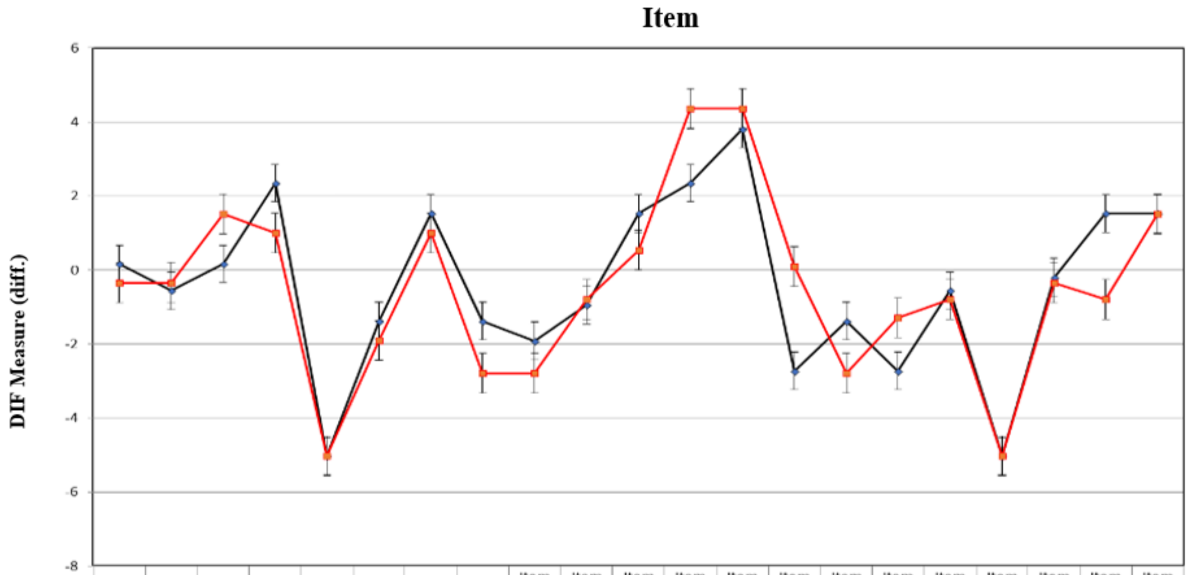
To find out how the Wright-Map works is to refer to locate of each item and person. In Figure 3 it is recognized that the 07P are the students who have the highest mathematical competence among the entire students. In contrast, 24P students are the ones with the lowest mathematical ability. Furthermore, Item No. 14 is an item that has the highest level of complexity among all the items, and vice versa item no. 6 is an item with the lowest level of complexity. Looking closely at the Wright-Map, the highest-ranking students (07P) were unable to completed 2 items (Item 14 and Item 13). On the other hand, 7 Items, i.e. No. 16, 10, 07, 17, 09, 20 and 6 can be exceeded well by all students.

Table 2. Distribution of standard competency based on item of mathematics test.

No.	Standards Competency	Items #
1.	Converts fractions to form percent and decimal and vice versa	1,6, and 11
2.	Sums and subtracts various fractional forms	2,7,12,17, and 19
3.	Multiply and divide various shapes	3, 8, 13, and 16
4.	Using fractions in comparison and scale issues	9, 14, 20, and 22
5.	Identify flat wake properties	5, 10, 15, 23, and 25

3.3. Impart of learning materials and gender differences on math

Based on Figure 3, the classroom teacher in the elementary school can confirm what items require to be improved in the student. For example, Item No. 5 is a problematic material because it is only able to be done by 1 student only (07P). In this context, teachers need to refine students' ability in learning to identify flat-wake characteristics. Concerning in the student side, Wright-Map also indicates which students need to become a special attention in learning math. For example, 26P, 27P, and 24P students who are students with the lowest learning ability among all students can be guided to create a group of mathematics learning.



	Item 1	Item 2	Item 3	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 19	Item 20	Item 22	Item 23	Item 25
L	0.1669	-0.562	0.1669	2.3492	-5.033	-1.379	1.5292	-1.379	-1.916	-0.947	1.5292	2.3492	3.8156	-2.731	-1.379	-2.731	-0.562	-5.033	-0.198	1.5292	1.5292
P	-0.341	-0.341	1.5131	0.9987	-5.033	-1.904	0.9987	-2.784	-2.784	-0.797	0.5355	4.3649	4.3656	0.0957	-2.784	-1.299	-0.797	-5.033	-0.341	-0.797	1.5131

Figure 4. The differences in ability between male and female students on each mathematical test item (measure in logits).

The average ability ratio between male and female students in mathematics is also not left from our investigation. In Figure 4 there is a difference in ability between male and female students in each item. The average male student has better math skills (-0.424 logits) than female students (-0.508). The difference ability on math between male and female students is also confirmed (+0.084 logits).

The findings in our study present a number of delightful knowledges. First, it is realized that the reliability of math tests is very satisfactory, but there are 4 test items that are questionable (Outfit). Outfit items indicate the non-functioning of items to differentiate, and measure students' abilities based on the pattern of responses given by the students [3, 23]. This can be reduced by fixing or eliminating the items. Besides that, there were also 7 students who were exposed to outfits. Students who outfit indicate if they experience problems when the exam is given. Problems during the exam can be in the form of students' misconceptions about the exam material, and/or students cheating during the exam. Secondly, Wright-Map clearly illustrates the ability of each student to work a math test. Students who have high abilities have a high grade of success in managing the exams [24]. At the same time, we can also see clearly the items displayed in Wright-Map also provide detailed information on which items have the highest to lowest complexity for students. Third, previous studies such as conducted by Nosek and Smyth [25], Dickhauser and Meyer [26], Bottia, Stearns, Mickelson, Moller and Valentino [27], Hyde and Mertz [28] mentioned there were differences in mathematical abilities between male and female students in several situations. Our findings also noticed the same phenomenon (Figure 4), where overall male students have better mathematical abilities than female students. However, on some items that measure we found female students have a mathematical ability rather than male students.

4. Conclusions

This study has shown the importance of applied a Wright-Map to assess student ability, specifically on math test. Wright-Map can map student performance in mastering the exam material provided. Used Wright-Map allows classroom teachers in elementary schools to estimate the quality of math learning in the classroom, and also covers any students who require assistance for supplementary learning. Another important finding in this study is male students have a stronger ability compared to female students in math exams. Furthermore, overall students have difficulties in fractional matter in comparative and scale problems, but they have no difficulty in operations to change fractions to form percent and decimal.

Acknowledgements

We thank to Mrs. Susilowati, S.Pd., for her permission to provide access to the results of student math exam at SDN Ceger 02 Jakarta Timur. No grants were accepted by the researchers in this study.

References

[1] Coombs C H 1952 A theory of psychological scaling *Engginering Research Bulletin* **34** 1-107
[2] Andrich D and Pedler P 2018 On a law of ordinal error. In: *2017 Joint IMEKO TC1-TC7-TC13 Symposium: Measurement Science Challenges in Natural and Social Sciences*: Institute of Physics Publishing)
[3] Sumintono B and Widhiarso W 2014 *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)* (Bandung: Trim Komunikata Publishing House)
[4] Wright B D 1977 Solving measurement problems with the Rasch model *Journal of Educational Measurement* **14** 97-116
[5] Engelhard Jr G 2013 *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*: Routledge)
[6] Sumintono B and Widhiarso W 2015 *Aplikasi pemodelan rasch pada assessment pendidikan* (Bandung: Trim Komunikata)
[7] Rangka I B, Prasetyaningtyas W E, Satrianta H and Folastri S 2017 Profil Perencanaan Karir Siswa Sekolah Menengah Kejuruan dengan Pemodelan Rasch Berdasarkan Jenis Kelamin *Konselor* **6** 39-48

[8] Che Musa N A, Mahmud Z and Baharun N 2017 Exploring students' perceived and actual ability in solving statistical problems based on Rasch measurement tools. In: *1st International Conference on Applied and Industrial Mathematics and Statistics 2017, ICoAIMS 2017*: Institute of Physics Publishing)
[9] Wilson M 2004 *Constructing measures: An item response modeling approach* (London: Lawrence Erlbaum Associates)
[10] Bond T G and Fox C M 2013 *Applying the Rasch model: Fundamental measurement in the human sciences*: Psychology Press)
[11] Wright B D and Stone M H 1979 *Best test design* (Chicago: Mesa Press)
[12] Wright B D and Masters G N 1982 *Rating scale analysis* (Chicago: Mesa Press)
[13] Hibbing A N and Rankin-Erickson J L 2003 A picture is worth a thousand words: Using visual images to improve comprehension for middle school struggling readers *The reading teacher* **56** 758-70
[14] Black P, Wilson M and Yao S Y 2011 Road maps for learning: A guide to the navigation of learning progressions *Meas.* **9** 71-123
[15] Boone W J 2016 Rasch analysis for instrument development: Why,when,and how? *CBE Life Sci. Educ.* **15**
[16] Bloom's B S 1965 *Bloom's taxonomy of educational objectives* (New York: Davis McKay Company)
[17] Dimitrov D M 2014 *Statistical methods for validation of assessment scale data in counseling and related fields*: John Wiley & Sons)
[18] Knoch U 2011 Investigating the effectiveness of individualized feedback to rating behavior — a longitudinal study *Language Testing* **28** 179-200
[19] Wilson M 2008 Cognitive diagnosis using item response models *J Zeitschrift für Psychologie/Journal of Psychology* **216** 74-88
[20] Andrich D 1988 *Rasch Models for Measurement* (Thousand Oaks, California: SAGE Publication)
[21] Linacre J M 2018 Winsteps® Rasch measurement computer program (Version 4.1.0). (Beaverton, Oregon: Winsteps.com)
[22] Rangka I B 2018 Datasets: Wright-Map to investigate the student's actual abilities in mathematics on class promotion test in elementary school. (USA: Center for Open Science)
[23] Smith R M and Suh K K 2003 Rasch fit statistics as a test of the invariance of item parameter estimates *J Appl Meas* **4** 153-63
[24] Engelhard G 2009 Using Item Response Theory and Model—Data Fit to Conceptualize Differential Item and Person Functioning for Students With Disabilities *Educational and Psychological Measurement* **69** 585-602
[25] Nosek B A and Smyth F L 2011 Implicit Social Cognitions Predict Sex Differences in Math Engagement and Achievement *American Educational Research Journal* **48** 1125-56
[26] Dickhauser O and Meyer W J P S 2006 Gender differences in young children's math ability attributions **48** 3
[27] Bottia M C, Stearns E, Mickelson R A, Moller S and Valentino L 2015 Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM *Economics of Education Review* **45** 14-27
[28] Hyde J S and Mertz J E 2009 Gender, culture, and mathematics performance *Proceedings of the National Academy of Sciences*