KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Paulina Skórska
Instytut Badań Edukacyjnych w Warszawie

Gender gap in reading and writing achievements

Streszczenie

RÓŻNICE OSIĄGNIĘĆ EDUKACYJNYCH DZIEWCZĄT I CHŁOPCÓW W ZAKRESIE CZYTANIA I PISANIA

Istniejące badania empiryczne wskazują, że dziewczynki uzyskują wyższe wyniki w testach czytania i pisania w języku ojczystym w porównaniu do chłopców. Różnice te wahają się od 0,15 do 0,6 odchylenia standardowego, a więc od niewielkiego do znaczącego. Wielkość tego efektu różni się także między krajami, etapami edukacyjnymi i latami. Celem tego badania było oszacowanie wielkości efektu różnic płciowych w wynikach polskich standaryzowanych testów o wysokiej doniosłości, w tym wypadku egzaminu gimnazjalnego z języka polskiego. W badaniu wykorzystano modelowanie *Item Response Theory*, a następnie oszacowano wielkość efektu różnic płciowych w wynikach, biorąc pod uwagę wrażliwość efektu na typ wykorzystanej statystyki. Wzięto pod uwagę różnice w wariancji wyników w grupie dziewcząt i chłopców, gdyż dotychczasowe badania często pomijały ten fakt, co mogło prowadzić do błędów interpretacyjnych. Wyniki wskazują na znaczący efekt różnic płciowych w wynikach z języka polskiego, większy niż można by się spodziewać na podstawie istniejących badań empirycznych, zarówno w zakresie średniej wielkości efektu, jak i różnic wśród uczniów o najwyższych i najniższych umiejętnościach (krańce rozkładu umiejętności). Rezultaty badania mogą być użyteczne dla nauczycieli i twórców edukacyjnych polityk publicznych, którzy są zainteresowani zapewnieniem realizacji zasady sprawiedliwości i równości w edukacji.

Słowa kluczowe: różnice osiągnięć chłopców i dziewcząt, umiejętności czytania i pisania, testy doniosłe, osiągnięcia uczniów, teoria odpowiedzi na pozycje testowe (*Item Response Theory*, *IRT*).

Introduction

In modern societies, education is one of the most important foundations of the whole system of social stratification and can be considered as one of the

channels of social mobility (Gromkowska-Melosik 2011). Therefore, fairness is one of the most fundamental issues in educational measurement as a basis for protecting both test takers and test users in all aspects of testing (AERA et al. 2014). The testing process is unfair if it does not represent certain individuals or groups in terms of the construct being measured (Kane 2010). In other words, a test is unfair if its results systematically underestimate or overestimate abilities of certain groups, e.g. boys and girls. Test fairness can be viewed as an important aspect of test validity (Messick 1989): any factors which reduce fairness also reduce measurement validity. Latest edition of Standards for Educational and Psychological Testing (AERA et al. 2014) underlines that fairness concerns the validity of individual scores interpretations for intended uses. In this approach, all members of a group, regardless of whether we are dealing with a girl or a boy, should be treated similarly when making interpretations of test scores for individuals.

Demarcation lines of access to various forms of socialization, as well as education and the labour market run along with gender (Gromkowska-Melosik 2011). Therefore, one of the primary tasks in the evaluation of test fairness according to gender differences is to separate the real differences in ability levels between boys and girls and the bias in the testing process and the test itself (gender as a factor generating construct-irrelevant variance). The realization of this objective requires, however, answering the question of whether gender gap in results of high-stakes tests still exists, as empirical studies indicate inconclusive evidences. This article focuses on gender gap in literacy tests, as many studies show that the gender gap in reading and writing abilities is still visible.

The majority of research concerning the ability to use native language relates to reading comprehension. The Progress in International Reading Literacy Study (PIRLS) indicated that 10-year-old girls tend to perform better in reading comprehension in 35 out of 40 countries (Mullis et al. 2007). In 2010 in all states girls performed better than boys in tests of reading ability (Chudowsky & Chudowsky 2010). Notwithstanding, the National Assessment of Educational Progress (NAEP) showed that gender differences in favour of girls decreased from 13 points (0.3 standard deviation) in 1971 to 7 points currently (0.15 deviation) (Rampey et al. 2009). However, empirical studies suggested different effect sizes of gender gap in reading ability. Sarah Logan and Rhona Johnston (2009) indicated that girls outperform boys in the reading tests of about 2/3 of the standard deviation, however Lietz (2006) argued slightly lower effect of gender gap amounting to 0.19 standard deviation. The Programme for International Student

Assessment (PISA) showed the difference of 0.3 standard deviations (OECD 2009) and PIRLS the gap of 0.2 standard deviation favouring girls (Robinson & Lubienski 2011). In 2010 edition of PISA, Polish students obtained results indicating that girls outperform boys by 0.36 standard deviation (Logan & Johnston 2010). Polish studies (Skórska & Świst 2014) indicated a gender gap of about 0.4 standard deviation favouring girls. At the same time, researchers pointed out a smaller gender gap in writing ability (Logan & Johnston 2009), which may reach about 0.04 of standard deviation and can be treated as negligible (Lietz 2006).

Some studies, however, showed very different results: there is no gender gap in reading ability or boys tend to slightly outperform girls (e.g. Knickerbocker 1989). The effect of gender gap in the reading comprehension can be sensitive to the culture: Johnson (1973–1974) pointed out that while in Canada and the USA girls outperform boys in reading, in England and Nigeria the effect was opposite. Studies showed that in some areas of literacy boys are better than girls, e.g. verbal analogies (Lietz 2006). Although it has been shown that reading gender gap exists in over 100 countries, the differences can be in large part explained by the different methodology of studies included in the meta-analyses and reviews (Lietz 2006). Recently Mikkel Wallentin (2009) presented a critique of a well-known meta-analysis computed by Janet Hyde and Marcia Linn (1988) and showed that if the effect sizes were weighed by group frequencies, the effect of gender gap would be opposite in direction, suggesting slightly better verbal skills of boys. In this meta-analysis, 66% of the studies indicated a negligible effect of gender differences. Wallentin (2009) noted that the biggest effects resulted from smaller studies (smaller sample sizes) – bigger studies suggested no effect at all.

The diversity of the results of the gender gap in literacy depends on the stage of education. Madhabi Chatterji (2006) showed negligible differences in reading ability in kindergarten and at the entrance to the primary school. Before the third year of primary school boys are achieving results comparable to girls, the first noticeable differences appear near to the 3rd grade (Husain & Millimet 2009). Some studies, however, showed that the variation in reading tests results may appear in the early stages of primary school (LoGerfo et al. 2006). At the beginning of primary school the effect equals to 0.12 standard deviation in favour of girls for the 90% of achievement distribution and ranges from 0.18 to 0.21 for the remaining part of the distribution. In 5th grade for top 25% of the reading test results the effect of gender gap becomes insignificant, and then again increases (Robinson & Lubienski 2011).

The evaluation of the gender gap in literacy, which is based only on the mean differences in reading and writing test results, can be misleading. Mean differences in test results between girls and boys may be still small, but the variances of the results in two groups may differ considerably. Large differences in test scores variances may translate into large differences in the results of boys and girls at both ends of the ability distribution – among the most and least talented students in the population. Variance of reading and writing ability for boys is typically higher than among girls by about 9% (Strand et al. 2006). Women have an advantage in each part of the reading ability distribution, but overrepresentation of boys is particularly visible at the left extreme of distribution, among the students with the worst test results (Robinson & Lubienski 2011).

Limitations of existing studies

Claudia Buchmann and colleagues (2008) emphasized that the debate on gender gap in literacy ability is not completed. As indicated previously, research on the gender differences between girls and boys are not conclusive. Many studies were based on mean performance differences between boys and girls, which is insufficient (Robinson & Lubienski 2011). In addition, the size of the effect clearly depends on the metric used. When the standardized results are used (as an indicator of the students achievements) gender gap is narrowing, whereas the analyses based on the scaled results indicate growth of the gap. Since, as noted above, the difference in the variances among boys and girls results is evident, the only proper metric used to comparison should be the effect size statistic. Moreover, most of the research was based on data collected in the Anglo-Saxon countries, especially on American student's samples. At the same time, patterns of gender inequalities in developing countries vary from those observed in the United States (Buchmann & Hannum 2001), so the results are not representative for other countries.

The aim of the study and methodology

The aim of the study was to examine gender gap in the results of large-scale, high-stakes tests in literacy in the Polish cultural context.

Student's achievements were estimated using results from Polish examinations after lower secondary school, from 2012 to 2014. In contrast to the majority of studies exploring gender gap in cognitive skills, in this study the data collected for the whole population (instead of just a sample) were used. The data were initially gathered by Central Examination Board. The size of population in the subsequent years was as follows:

- 393,766 students (49.19% girls and 50.81% boys) in 2012;
- 379,752 students (48.97% girls and 51.03% boys) in 2013;
- 362,755 students (48.96% girls and 51.04% boys) in 2014.

Student's ability level was estimated by using standardized total test scores and scaled scores (with Item Response Theory (IRT) modeling). Thereby, it was possible to compare the magnitude of the gender gap estimated in two different ways, since, as stated above, the size of the gender gap effect is sensitive to the metric used.

IRT relies on two basic assumptions (Hambleton, Swaminathan & Rogers 1991): a) the performance of a student on a particular test can be predicted by a set of latent traits (abilities) and b) the relationship between students' item performance and these latent traits can be described by the monotonically increasing function called item characteristic curve (ICC). This function indicates that for students with higher levels of latent trait (ability), the probability of the correct response to particular item increases.

The three most popular IRT models take their names from the number of item parameters (characteristics) estimated in the model. One-parameter logistic model (1PLM) is often called the Rasch model (Rasch 1960). In 1PLM only one item parameter – the difficulty (b parameter) is estimated. The more difficult particular item is, the higher level of latent trait (ability) is needed to answer this item correctly. The difficulty is also called location parameter, because it indicates the location of ICC relatively to ability scale. Lower b-values are associated with shift of ICC to the left part of the ability scale, suggesting that an item is easy. Two-parameter logistic model (2PLM; Birnbaum 1968) estimates also discrimination (a) parameter (in 1PLM model, for all items constant, equals to 1). Discrimination parameter indicates the slope of the ICC (how steep the ICC is for a particular item). The higher the discrimination parameter is, the better differentiation between students with different ability levels. Three-parameter logistic model (3PLM; Samejima 1969) allows for non-zero lower asymptote of ICC, and therefore the third parameter called the guessing or pseudo-guessing (c) is estimated. C parameter represents the probability of providing the correct

answer in the item by students with a very low ability level. Other models in IRT framework are the extensions of these three fundamental ones[1].

In this study, for the dichotomous items three-parameter logistic model (3PLM) was used to estimate students' reading and writing ability. For the vast majority of items 3PLM exhibited a good fit to the data. For the rest of items two-parameter logistic model (2PLM) was used. For polytomously scored items Graded Response Model (GRM; Samejima 1969) was used to estimate students' ability level and item parameters.

As students' ability level (θ) is computed on the basis of only a subset of the total possible item pool, it is estimated with a substantial amount of measurement error. To take this uncertainty associated with the estimates into account, and to compute unbiased group-level estimates, plausible values (PV) were used. Plausible values are multiple values representing the likely distribution of a student's ability and are generated through multiple imputations (Mislevy 1993). In this study five plausible values per individual were estimated.

To estimate the effect size of gender gap in reading and writing ability, procedures traditionally occurring in meta-analysis (MA) were used. The effect of gender gap was estimated using standardized mean difference (Hedges & Olkin 1985). To assess discrepancy of literacy achievements between boys and girls, variance ratio (VR) was used. This is the ratio of the male variance to the female variance (Feingold 1992). Thus, a VR greater than 1.00 indicates greater male variability, VR of 1.00 indicates homogeneity of variance across gender and VR less than 1.00 indicates greater female variability.

The gender gap on both ends of ability distribution (among the most and least talented students in the population) was defined as a ratio of girls to boys (and reverse) for the lowest and highest 10, 5 and 1 percentage of ability estimates.

Results

The effect size of gender gap in literacy achievements, measured as a standardized mean difference between girls and boys, is generally slightly higher for IRT ability estimates than for other metrics. For IRT estimates, standardized differences in means of girls and boys amount to 0.47, 0.46 and

[1] More can be read in e.g. Kondratek and Pokropek (2013).

0.46 in 2012, 2013 and 2014, respectively. For plausible values the gender gap effect size is slightly smaller, ranging from 0.42 (in 2014) to 0.44 (in 2012). Details about the effect sizes and their corresponding confidence intervals (CI) can be found in Table 1.

Table 1. Effect sizes of gender gap in literacy tests (examination after lower secondary school) 2012–14 with corresponding confidence intervals according to the metric used

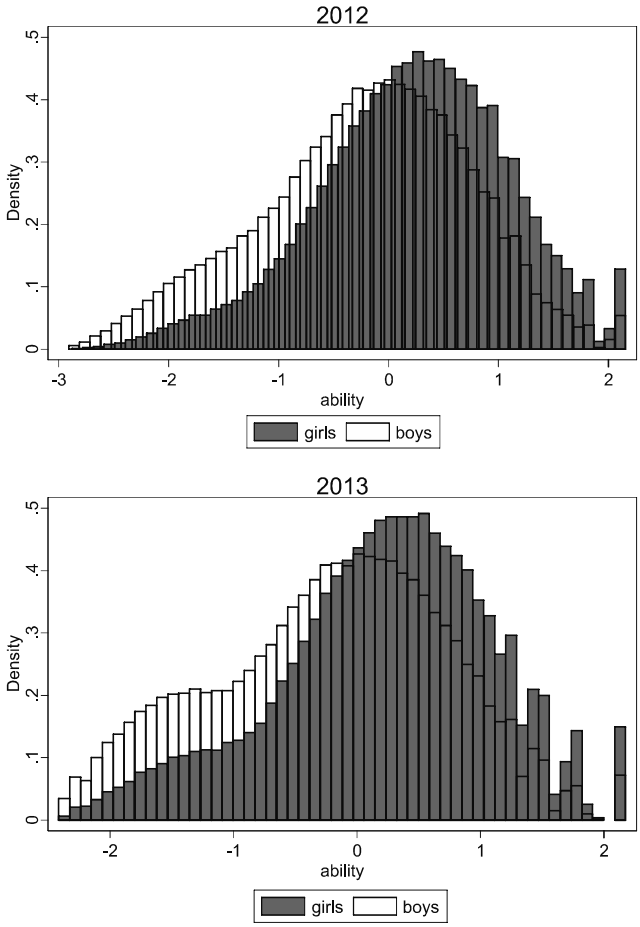
Year	Estimate	Effect size	Lower bound of CI	Upper bound of CI
2012	IRT (θ)	0.47	0.46	0.47
	PVs	0.44	0.43	0.44
	standardized total scores	0.45	0.44	0.46
2013	IRT (θ)	0.46	0.45	0.46
	PVs	0.43	0.42	0.43
	standardized total scores	0.42	0.42	0.43
2014	IRT (θ)	0.46	0.45	0.46
	PVs	0.42	0.42	0.42
	standardized total scores	0.43	0.42	0.43

* Final estimates of effect size for plausible values were computed by averaging effect sizes for five PVs.

As noted above, effect size defined as standardized mean difference is insufficient, since relying on average performance may translate into attenuation of gender gap effect magnitude. Examination of Graph 1 reveals evident differences between boys and girls in performance (estimated as θ in IRT models) on literacy test virtually for the entire ability distribution. Boys are overrepresented for the lower levels of ability (left part of the distribution) whereas girls dominate in the right part of the distribution.

As can be seen in Table 2, boys' ability estimates, regardless of whether they are specified as scaled (with IRT models) or standardized scores, are more variable than girls results for each year concerned. Taking into consideration IRT ability estimates, the variance ratio (VR) is the highest for 2012 examination and lowest in 2014. In percentage terms boys' results are 19% more variable than girls' in 2012 and 18% more variable in 2013. The magnitude of variance ratio is, as expected, sensitive to the metric used to estimate literacy ability, as standardized total test scores indicate the highest and PVs the lowest values of VR. According to standardized total test scores differences in ability variance among

boys and girls are meaningful, as boys' scores are even 23% more variable than girls (in 2012).

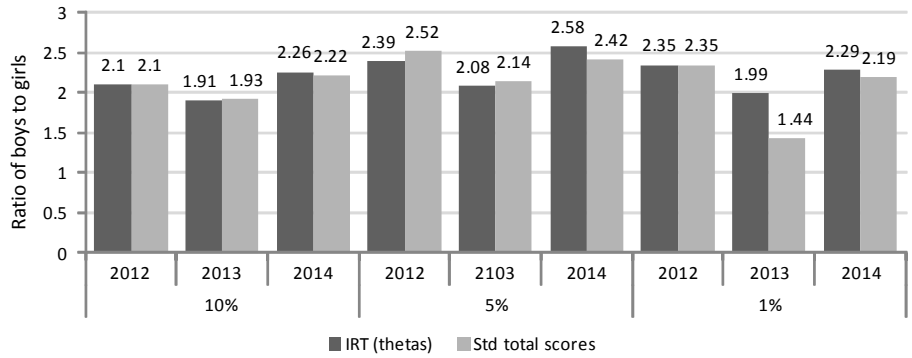


Graph 1. Distributions of literacy ability (estimated as θ in IRT models) 2012–14 according to gender

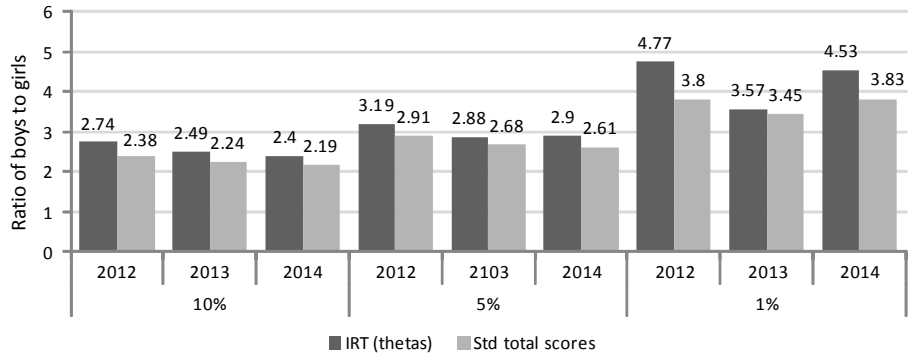
Table 2. Variance ratios (VRs) for the results of examination after lower secondary school in 2012–14 according to the metric used to estimate literacy ability

Year	IRT (θ)	PVs	Standardized total scores
2012	1.19	1.16	1.23
2013	1.18	1.16	1.19
2014	1.10	1.07	1.22

High differences in variance between boys and girls results may translate into considerable gender gap for the most and least talented parts of population and this hypothesis was examined. The ratios of girls' to boys' scoring above 90% are from 1.91 in 2013 to 2.26 in 2014 (for IRT ability estimates). The biggest gender gap can be observed in 2013, when 2.5 times more girls than boys achieved scores above 95% of distribution. Using standardized total test scores as ability estimates, similar results can be observed (except 2013), indicating twice the number of female among the most talented students in the population.



Graph 2. Ratios of girls to boys in 10, 5 and 1 percentage of the highest literacy ability estimates (θ in IRT models and standardized total test scores) in 2012–14



Graph 3. Ratios of boys to girls in 10, 5 and 1 percentage of the lowest literacy ability estimates (θ in IRT models and standardized total test scores) in 2012–14

As can be seen in Graph 3, even more evident gender inequalities can be observed at the lower tail of the ability distribution. For the 1% of the least talented students the overrepresentation of boys is clear. In 2012 in the group of lowest scoring students, there were nearly five times more boys than girls.

In 2013 this ratio was lower, but in 2014 once again exceeded 4.5. For both, the upper and lower tail of ability distribution, disparities in the number of girls and boys are most pronounced for 1% of the most and least talented students in the population.

Discussion

The study confirmed that the gender gap is still visible in the literacy ability for students at the end of the lower secondary school. As expected (Robinson & Lubienski 2011), using scaled scores as estimates of literacy ability produces slightly higher gender gap effect sizes in comparison to standardized scores. Regardless of the metric used to estimate literacy ability, according to the rule of thumb proposed by Jacob Cohen (1992), the magnitude of gender gap can be classified as moderate. In practical terms effect sizes obtained in this study indicate that the mean literacy score of girls is at the 70% of boys' literacy scores. Translating this gap into common language effect size, also known as the probability of superiority (Grissom & Kim 2005), leads to the conclusion that the probability that a randomly sampled girl will have a higher literacy ability level than a randomly sampled boy is about 33%. In terms of differences in variability, the distribution of scores for the girls overlaps only in 2/3 with the distribution of scores for boys, and nonoverlap amounts to 33%. This lack of overlap can be confirmed by visual examination of both distributions. Gender differences can be observed essentially in any part of the ability distribution, especially for both lower and upper tail.

The effect of gender gap is considerable and slightly higher (difference about half a standard deviation) than reported in Anglo-Saxon countries. This suggests that gender gap in reading and writing is strongly sensitive to the culture, as indicated by Johnson (1973–1974). The magnitude of the gender gap in education is strongly dependent on the socio-economic conditions in the particular country. Luigi Guiso and colleagues (2008) evidenced that gender differences in school achievements are related to Gender Gap Index (GGI). Gender Gap Index consists of four dimensions: health, educational attainment, economic participation and opportunity, and political empowerment. Thus, it reflects the global gender inequalities in key areas of social life. According to The Global Gender Gap Report (2013) results, the rank of Poland decreased from 44. in 2006 to 54. in 2013, indicating growing gender inequalities, relative to other 135 analyzed countries.

Taking into consideration these results, further studies should examine whether the gender gap in reading and writing skills widens over time.

Differences in variances of ability estimates in a group of girls and boys are meaningful, especially for IRT estimates and standardized total test scores, where the variance ratio suggests that boys' scores are about 20% more variable than girls' scores. These results are consistent with the trends found in other countries, however the variance ratio (VR) magnitude is about twice as high as VR reported in the other empirical studies (e.g. Strand et al. 2006).

Such big differences in variances of girls and boys performance may translate into considerable gender gap for the most and least talented parts of population. This hypothesis found a confirmation in this study. The highest gender gap can be observed for the lower tail of literacy ability distribution. Among the students with lowest 1 percentage of ability scores, there is over four times more boys than girls. In 2012 boys made up more than 80% of students in this part of distribution. This effect is higher than should be expected based on existing studies concerning assessment of gender gap in high-stakes tests (e.g. Hyde et al. 2008).

The possible explanations of gender gap in the literacy can be classified into a few categories. Some scientists underline that boys and girls differ in reading and writing achievements due to biological differences (Gunzelmann & Connell 2006). Left brain hemisphere, responsible for verbal activities, develops earlier for girls than for boys. Males' fine motor also develops later, leading to difficulties in mastering the mechanic activities needed for reading and writing (e.g. holding a pen). Another set of potential explanations is associated with socialization. Some studies (e.g. Wilhelm & Smith 2009) showed that boys may consider reading as a "feminine" activity. Therefore some students may reject reading, trying to confirm male identity, which may reduce boys' motivation to reading and lead to a self-fulfilling prophecy. Boys and girls also differ in learning strategies. Girls apply more strategies and use meta-knowledge, whereas boys are focused on information retrieval, preferring rapidity of responses over deep understanding of text (Smith & Wilhelm 2002). In recent years, PISA study showed (OECD 2009) that reading engagement is one of the best predictor of literacy achievements, sometimes even stronger than socioeconomic status. Boys are less engaged in reading than girls and moreover, among boys reading comprehension is affected by the content of text. The question whether gender gap in literacy among Polish students is caused more by biological, social or psychological factors should be the subject of further research and analysis.

The study is exploratory and only shed little light on the description and magnitude of gender gap in literacy achievements, measured by high-stakes tests (examination after lower secondary school). The possibility to assess the change of gender gap over time is limited, as the results include the data form examination in three sequential years. The conclusion based on analyzed data is that gender gap does not tend to narrow between 2012 and 2014, either in terms of mean effect size, or in terms of frequency of boys and girls among the most and least talented students. Considering the nature of effect size statistic, which is in principle a standardized metric, the comparison of gender gap over time is methodologically proper, however using equated test scores will increase results validity. Further research will be focused on the assessment of gender gap in other educational stages (primary and upper-secondary school), as well as changes in phenomenon over time, using Comparable Examination Results (CER)[2].

References

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. 2014. *Standards for educational and psychological testing*, American Educational Research Association, Washington.

Birnbaum A. 1968. Some latent trait models and their use in inferring an examinee's ability, in: Lord F.M. & Novick M.R. (eds.), *Statistical theories of mental test scores*, Addison-Wesley, Reading, pp. 17–20.

Buchmann C., DiPrete T.A. & McDaniel A. 2008. Gender inequalities in education, *Annual Review of Sociology*, *34*, pp. 319–337.

Buchmann C. & Hannum E. 2001. Education and stratification in developing countries: A review of theories and empirical research, *Annual Review of Sociology*, *27*, pp. 77–102.

Chatterji M. 2006. Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the early childhood longitudinal study (ECLS) kindergarten to first grade sample, *Journal of Educational Psychology*, *98*(3), pp. 489–507.

Chudowsky N. & Chudowsky V. 2010. *Are there differences in achievement between boys and girls?*, Center on Education Policy, Washington, DC.

Cohen J. 1992. A power primer, *Psychological Bulletin*, *112*, pp. 155–159.

Feingold A. 1992. Sex differences in variability in intellectual abilities: A new look at an old controversy, *Review of Educational Research*, *62*, pp. 61–84.

Grissom R.J. & Kim J.J. 2005. *Effect Sizes for Research: A Broad Practical Approach*, Lawrence Erlbaum Associates, Mahwah, NJ.

[2] More about CER can be found on: http://pwe.ibe.edu.pl

Gromkowska-Melosik A. 2011. *Edukacja i (nie)równość społeczna kobiet. Studium dynamiki dostępu*, Oficyna Wydawnicza Impuls, Kraków.

Guiso L., Monte F., Sapienza P. & Zingales L. 2008. Culture, Gender and Math, *Science*, *320*(5880), pp. 1164–1165.

Gunzelmann B., & Connell D. 2006. The new gender gap: Psychological, neurobiological, and educational perspectives, *Educational Horizons*, *84*(2), pp. 94–101.

Hambleton R.K., Swaminathan H. & Rogers H.J. 1991. *Fundamentals of Item Response Theory*, Sage, Newbury Park.

Hedges L.V. & Olkin I. 1985. *Statistical methods for meta-analysis*, Academic Press, Orlando.

Husain M. & Millimet D. 2009. The mythical "boy crisis"?, *Economics of Education Review*, *28*, pp. 38–48.

Hyde J.S. & Linn M.C. 1988. Gender differences in verbal ability: A meta-analysis, *Psychological Bulletin*, *104*, pp. 53–69.

Johnson D.D. 1973–1974. Sex differences in reading across cultures, *Reading Research Quarterly*, *9*(1), pp. 67–86.

Kane M. 2010. Validity and fairness, *Language Testing*, *27*(2), pp. 177–182.

Knickerbocker J.L. 1989. Sex differences in reading achievement: A review of recent research, *Ohio Reading Teacher*, *24*(2), pp. 33–42.

Kondratek B. & Pokropek A. 2013. IRT i pomiar edukacyjny, *Edukacja*, *124*(4), pp. 42–66.

Lietz P. 2006. A meta-analysis of gender differences in reading achievement at the secondary school level, *Studies in Educational Evaluation*, *32*, pp. 317–344.

Logan S. & Johnston R.S. 2009. Gender differences in reading ability and attitudes: Examining where these differences lie, *Journal of Research in Reading*, *32*(2), pp. 199–214.

Logan S. & Johnston R.S. 2010. Investigating gender differences in reading, *Educational Review*, *62*(2), pp. 175–187.

LoGerfo L., Nichols A. & Reardon S.F. 2006. *Achievement gains in elementary and high school*, Urban Institute, Washington, DC.

Messick S. 1989. Validity, in: Linn R.L. (ed.), *Educational measurement*, American Council on Education, Macmillan, New York, pp. 13–103.

Mislevy R.J. 1993. Should "multiple imputations" be treated as "multiple indicators"?, *Psychometrika*, *58*(1), pp. 79–85.

Mullis I.V.S., Martin M.O., Kennedy A.M. & Foy P. 2007. *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*, Boston College, Chestnut Hill, MA.

Organisation for Economic Co-operation and Development. 2009. *Equally prepared for life? How 15-year-old boys and girls perform in school*, Organisation for Economic Co-operation and Development, Paris.

Rampey B.D., Dion G.S. & Donahue P.L. 2009. *NAEP 2008 trends in academic progress* (NCES 2009-479), National Center for Education Statistics, Washington, DC.

Rasch G. 1960. *Probabilistic models for some intelligence and attainment tests*, Danish Institute for Educational Research, Copenhagen.

Robinson J.P. & Lubienski S.T. 2011. The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings, *American Educational Research Journal*, *48*(2), pp. 268–302.

Samejima F. 1969. *Estimation of latent ability using a response pattern of graded scores*, Psychometrika monograph (Vol. 17).

Skórska P. & Świst K. 2014. Wielkość efektu płci w wewnątrzszkolnych i zewnątrzszkolnych wskaźnikach osiągnięć ucznia, [in:] Niemierko B. & Szmigiel M.K. (eds.), *Diagnozy Edukacyjne. Dorobek i nowe zadania*, Grupa Tomami, Kraków, pp. 89–103.

Smith M.W. & Wilhelm J.D. 2002. *Reading don't fix no chevys: Literacy in the lives of young men*, Heinemann, Portsmouth, NH.

Strand S., Deary I.J. & Smith P. 2006. Sex differences in Cognitive Abilities Test scores: A UK National Picture, *British Journal of Educational Psychology*, *76*, pp. 463–480.

Wallentin M. 2009. Putative sex differences in verbal abilities and language cortex: A critical review, *Brain and Language*, *108*, pp. 175–183.

Wilhelm J. & Smith M.W. 2009. Boys and literacy: Complexity and multiplicity, in: Christenbury L., Bomer R. & Smagorinsky P. (eds.), *Handbook of adolescent literacy research*, Guilford Press, New York, pp. 360–371.

World Economic Forum. 2013. *The Global Gender Gap Report* (2013). Available: http://www3.weforum.org/docs/WEF_GenderGap_Report_201.