ON THE SEARCH FOR GENDER-RELATED DIFFERENCES IN DUTCH PRIMARY MATHEMATICS CLASSROOMS

Marja van den Heuvel-Panhuizen

Freudenthal Institute, Utrecht University

This paper reports on a study carried out in the Netherlands which is aimed at finding an explanation for the curious evidence that the reformed approach to mathematics education appears to have a less positive influence on girls' results. Over and over it turns out that girls' achievement scores in primary mathematics are lower than those of boys. The research started with 5000 schools and ended up with 4 schools. The paper focuses on the last part of this zooming-in research in which the classroom observations took place. By means of these observations a number of classroom characteristics, which make the teaching less optimized for girls and may cause their mathematics scores to be lower, have been traced.

INTRODUCTION

Since 1987, the National Institute of Educational Measurement (Cito) investigates the mathematics achievements of Dutch primary school students on a national scale. This takes place with a five year interval. As well as making visible the developments in the yield of education, these so-called PPON studies, have served to emphasize gender differences in mathematics scores. It turns out that boys are systematically outperforming girls. See the PPON results reported by Wijnstra (1988), Bokhove et al. (1996) and Jansen et al. (1999). It is unclear how long these differences have existed. Research (Wijnstra, 1982) had already indicated in the early 1980s that girls were behind in mathematics in primary schools in the Netherlands. Internationally the situation is different. Although the findings of international research are not always unequivocal, on the whole the situation is that the further along one moves in education, the more the boys outperform the girls. In general, no differences are found in primary education, and if they are, they are usually to the advantage of girls (e.g., Hyde et al., 1990; Leder, 1992; Geary, 1994). In addition the latest development is that in several countries girls are starting to do better in secondary education as well (e.g., Shaw, 2002). However, this is not the case in the Netherlands. The atypicality of our results was proven yet again in TIMSS (Mullis et al., 1997). For example, significant gender differences in grade 4 were only found in the Netherlands, Japan and Korea.

Girls' lagging result lead to the intriguing question of whether the Dutch approach to mathematics education, called "Realistic Mathematics Education" (RME), is equally suited for both girls and boys.

The foundations for the RME approach were laid by Freudenthal and his colleagues in the early 1970s. A brief overview of the philosophy and principles of RME can be found in Van den Heuvel-Panhuizen (2001). The significance of RME lies in its focus on mathematics that is worthwhile to learn and makes sense to the students. RME tries to achieve these goals by making mathematics experientially real for the students and having them actively involved in the learning process. In short, the RME approach means that mathematics education starts with rich context problems that can be solved in different ways. By means of interactive classroom discussions and the use of models, the inital context-connected strategies gradually evolve to more general, formal solutions reflecting a higher level of understanding.

THE MOOJ STUDY

Although there had been indications for a long time that girls did less well in mathematics in primary school than boys, it was not until the mid 1990s that research was done into these gender differences. In 1995, the Freudenthal Institute of Utrecht University and the Center for Study of Education and Instruction of the State University of Leiden received a grant from the Dutch Ministry of Education to start this research. It resulted in a collaborative project in which Cito was involved as well. The study was called the MOOJ study, and lasted until 1999. For a detailed report, see Van den Heuvel-Panhuizen & Vermeer (1999).

The purpose of the MOOJ study was twofold: (1) getting an overview of the size and nature of gender differences in mathematics achievements in primary school; (2) finding mechanisms in mathematics classrooms that contribute to these differences. The way the MOOJ study was set up was rather different from the customary design in educational research. The study consisted of three phases, with Stage I starting with around 5000 schools, while Stage II zoomed in on 14 schools and Stage III zoomed in even further on only 4 schools. The focus in the present research paper is on this final part of the study. Before going into this in more detail, first a short summary of the first two parts.

Stage I was meant mainly to map gender differentiation and to identify schools where the differences were higher or lower. To accomplish this, the mathematics scores in the Cito End of Primary School Test in 1993, 1994 and 1995 were analysed for gender differences. In these years, this test was administered in about 70% of the sixth-grade classes (the students are 12 years old at that point) , which led to a data base of mathematics scores from approximately 100,000 students. The results that were gained from the analysis of these scores were presented at PME 21; see Van den Heuvel-Panhuizen, 1997). In total, gender differences at three analysis levels were found. The first finding was that in each of the three years the avarage total scores of boys were about 6% points higher than the avarage total scores of the girls. This difference is about a quarter of the standard deviation. The second finding was that the test items showed remarkable gender-specific characteristics. Particular problems (called “boys problems”) were always done better by the boys and some other problems were done relatively well by the girls (called “girls problems”). The third finding was that in half of the schools the boys outperformed the girls (these schools were called “boys schools” or “B-schools”). In the other schools the average score of the girls was equal to that of the boys or higher (these schools were called “girls schools” or”G-schools”).

In Stage II, the study zoomed in to 7 B-schools and 7 G-schools to collect additional information about teachers and students. One of the main findings of this part of the study was that there were strategy differences between boys and girls. (For a more detailed report, see Van den Heuvel-Panhuizen & Vermeer, 1999).

MOOJ STUDY _ STAGE III

Stage III was the most crucial part of the MOOJ study. In this part, observations took place in 4 sixth-grade classes to see if any specific patterns exist that may explain the gender differences in mathematics achievement. The observations took place in April 1997 in two B-schools and two G-schools. These schools were selected from the Stage I and II data. They were "extreme" schools, which means that in the B-schools the differences were most to the boys' advantage and in the G-schools the differences were to the girls' advantage or were smallest. The rationale for selecting extreme schools is that in these schools the chance of observing gender differences and related factors is largest. Such a selection is in agreement with the idea of purposeful sampling that is characteristic for the "Grounded Theory" approach as formulated by Glaser and Strauss (1967) and others.

To answer the question of why the girls get the same score as boys in particular schools and why they do not in other schools four lessons were observed in each of the four grade 6 classes. The teachers were free to choose a particular content for these lessons (with the exception of the fourth lesson) and were not informed about the gender-specific aim of the observations. In the request to the teachers to be allowed to make the observations in their classrooms, they were told that the researchers wanted to gain knowledge about classroom practice. The observations had a double-focused set up. During each lesson observations were made from two different perspectives. The Leiden team had a general didactical perspective and collected data about general aspects of verbal communication within the classroom. The Utrecht team observed from a domain-specific didactical perspective and focused mainly on the characteristics of the learning situations that occurred within the lessons. After both sets of observations were finished, the findings of the two teams have been linked to each other. For analysis of the data, use has been made of the "Constant Comparative Method" (Glaser & Strauss, 1967; Strauss & Corbin, 1990) which implies repeatedly moving back and forth between the data that were found in the different classes and by the different observers, the observers thoughts about the data, and the conjectures made about it, finally resulting in some conclusions with which the whole team could agree.

The domain-specific didactical perspective

The observations of the Utrecht team were focused on mathematical content and the teaching methods in the lessons. The observations were carried out by four experienced mathematics educators and researchers Adri Treffers, Leen Streefland, Koeno Gravemeijer and the author of this paper who also prepared the observation format.

The theoretical base that was taken as the starting point for the development of the observation and analysis points consisted of:

- *Didactical characteristics of RME*; such as offering learning opportunities by (a) paying attention to different strategies and their relation, (b) developing number benchmarks, (c) developing knowledge about daily-life measures, (d) developing estimation strategies.
- *Gender-specific interaction characteristics* (Jungwirth, 1991, 1996) that are related to (I) determining competence such as (a) undoing completeness vs. teacher's echo, (b) authority insistence vs. argumentative insistence, (c) emerging failure vs. concealing of failure; or that are related to (II) learning opportunities such as (a) blocking task constitution vs. tasks constitution, (b) blocking outside reference vs. demonstrating everyday knowledge.

- *Class climate characteristics* (e.g. Cobb, Wood & Yackel, 1991) related to (a) learning (what are the social norms regarding responsibility and autonomy of students?); (b) subject matter (what are the socio-math norms regarding e.g. different strategies, estimations, real world references); and (c) evaluation (what are the social norms and the socio-math norms regarding who is determining what is correct/incorrect?).

The procedure for the observations was as follows. The four observers each attended one lesson of each of the four teachers. During the lessons they made notes about significant episodes based on the above list of characteristics that they were free to apply based on their own insights. After observing the lesson they made a report consisting of three parts: (1) a description of the lesson, (2) theoretical memos about what they saw in the lesson, and (3) a prediction of the classroom type (B-classroom or G-classroom) including the arguments this was based on.

With the exception of the author who was the project leader of the MOOJ study and had to make the arrangments and schedule for the observations, the observers were not aware about the classroom type.

After all observations had taken place, the observers studied each other's reports, and after a short intermezzo to let everything sink in, a meeting was held during which the observers reacted to their respective findings, and where finally conclusions, with which all four observers concurred, were formulated.

The general didactical perspective

For the observaties by the Leiden team use has been made of the FROG tool (Dolle-Willemsen, 1997). This is a computer-based observation tool for classroom interaction that covers several categories of classroom activities. The most important categories are: explaining or demonstrating by teacher, asking questions by teacher, pausing, giving turns to boy/girl, answering questions by boy/girl, taking initiative by boy/girl, reacting by teacher. The computer screen shows the categories and the observer has to score the category each time the category changes. The lessons were all scored by one observer who had a large experience with this tool, and who also was not aware of whether the lesson was taking place in a G-classroom or a B-classroom. For the analysis of the scores, both the frequency and the amount of time spent on specific categories was taken into account.

RESULTS CLASSROOM OBSERVATIONS

Because it is impossible to dicuss all results in this paper, they will only be summarized here.

Results from the Utrecht observations

One of the points to come out of the Utrecht observations was that regarding *the class climate characteristics*, the following points were seen as being positive for girls' learning achievements: security, mutual respect and an ordered atmosphere with clear social rules. The classrooms that fulfilled these characteristics were nearly always characterized as G-classrooms.

As can be seen in Table 1, the predictions of the *classroom type* almost completely agree with the actual nature of the classroom. In total ten out of twelve possible predictions

	Predictions of the classroom type and the observers' arguments

	Classroom #2 (G)	Classroom #7 (G)	Classroom #13 (B)	Classroom #6 (B)
Observer U1	--	**G-Classroom** –less uncertainty –no challenge –no reason to be afraid of mistakes –focused on routines	**probably B-Classroom** –insufficiently safe atmosphere –strong focus on smartest solution –ego-focused –more participation boys	**probably G-Classroom** –focus on procedures –focus on CITO-like abilities –focus on achievement (= disadv. girls) –ego-focused (= disadv. girls)
Observer U2	**G-Classroom** – knowing strategies in advance –large degree of clarity in organisation –large degree of clarity in assignments –honest teacher's report that strategies were emphasized because of the research	**G-Classroom** –good atmosphere –good organization –structured way of working –boys made the mistakes	**B-Classroom** –instruction is rule-directed, but the speed is high –the lack of sufficient learning opportunities –problems with concentration (mixed-grade class) –textbook series with few possibilities to apply routines –didactical deficits of the teacher –more questions from boys	**B-Classroom** –high speed of the instruction –verbal character of explanations
Observer U3	**clear G-Classroom** –quiet atmosphere –structured teaching –teacher shows respect –didactics aimed at rules overall impression: affective + cognitive +/–	**clear G-Classroom** –good atmosphere, paternal teacher shows respect –didactics aimed at rules overall impression: affective + cognitive +/–	**B-Classroom** –unordered, chaotic–rather impersonal approach – very little mutual respect overall impression: affective – cognitive +/–	**B-Classroom** –aimed at indidual work; not enough collective moments –not enough learning support –students addressed too much on their individual qualities –shortcomings in didactics overal impression: affective +/+,– cognitive –
Observer U4	**G-Classroom** –secure and orderly classroom atmosphere –a lot of social room –instrumental explanation –no clear appeal to individual input (=indirect adv. girls) –leaving learning opportunities unused (=disadv. girls) –little initiative from girls (= disadv. girls)	**G-Classroom** –safe social atmosphere; teacher respects the students –didactically safe atmosphere; structured didactics and fixed approach –little emphasis on initiative, reflection, critical attitude (=indirect adv. girls)	**B-Classroom** –complex organisation; order problems and inefficiency in explanations–limited didactic quality –influence of realistic method (certain sums, models and approaches) that boys may be gaining more from	**B-Classroom** –not a really secure atmosphere –instrumental expalanation, but failing didactics (for example in estimation) –tasks are ambivalent –not always understanding children and not really looking into what they have found

Table 1: Summary of the observers' predictions of the classroom type (G or B) and their arguments

were correct (twelve predictions could be made, the four made by Observer U4 do not count, since this observer was aware of the type of classroom), Although this high correlation needs to be relativized because of the low number of degrees of freedom, especially the agreement in argumentation gives good indications for answering the research question.Based on the observations from the Utrecht team, few to no conclusions could be drawn regarding the *interaction characteristics.* This was caused in the first place by Jungwirth's interaction characteristics only being recognized in a very limited way during the lessons. Furthermore there was no gender specific pattern. See, for example, the following two observation, both from classroom #7, a G-classroom, and both relating to a boy.

Undoing completeness:	[Classroom #7; Lesson 1; U4] Harry gives a good answer to an exercise straightaway. The teacher did not expect this and starts to explain to the whole class how to get to this answer. (By explaining so elaborately how Harry got this answer, the suggestion is made that Harry himself cannot do this.)
Teacher's echo:	[Classroom #7; Lesson 1; U4] The teacher corrects Fred's answer and adds that Fred made a mistake. (Because the teacher says that Fred only made a slip of the tongue, he in facts indicates that Fred is competent.)

Despite the fact that Jungwirth's gender-specific characteristics were only found to a certain degree, it became clear that they can expose interesting mechanisms which can have an unmistakable influence on mathematics achievements, for example, by having an effect on the arising or not arising of learning opportunities (e.g. blocking task constitution vs task constitution) and by influencing how both students themselves and others regard their competence (e.g. undoing completeness vs teacher's echo).

When taking *didactical characteristics* as the point of view, the shortfalls in the implementation of realistic didactics stood out especially (a more detailed look at this, based on classroom vignettes, will be taken in the presentation at the conference). Classrooms where didactics fell short according to the observers were often classified as B-classrooms. Classrooms with a lot of structure and instrumental explanations were often classified as G-classrooms.

In short, the observations of the Utrecht team contain indications that a socially and cognitively secure atmosphere worked immediately to the advantage of the girls. An additional indirect advantage for the girls was that in the G-classrooms own cognitive input (such as knowledge of measures) and social input (for instance, taking the initiative) were not expected.. This was called an indirect advantage because it meant that boys could distinguish themselves less.. In the B-classrooms on the other hand, the lack of a clearly secure atmosphere worked directly to the disadvantage of the girls. Indirect disadvantages for the girls here were didactic shortcomings and insufficient learning opportunities (for example teaching wrong strategies, dismissing correct solutions, not teaching how to estimate, not building knowledge of measurements). More and more the perception arose that a bad implementation of RME is more disadvantageous for girls than for boys.

Results from the Leiden observations

It became apparent from the analysis of the Leiden team observation data that more thinking questions were asked in the two G-classrooms and that there were more breaks to think than in the B-classrooms. Especially classroom #7 devoted a relatively large amount of time to thinking breaks. This fits the pattern of cognitive support that was recognized, on the basis of the Utrecht observations, as an important characteristic of the observed lessons in this classroom. Furthermore, a more detailed analysis of the frequency of the categories showed that on the whole the students were taking a more active role in the learning process in the two G-classrooms. This point also arises from the observation that generally speaking more questions were asked in the G-classrooms.

DISCUSSION

The context of the research question at the basis of the MOOJ study is very complex and giving *the* final answer is difficult. The research has not only resulted in many new research questions, but also in a number of very penetrating points for discussion; especially regarding the role of RME. The RME approach to learning may be better teaching than the traditional education that existed in the Netherlands twenty-five years ago, but measured by mathematics achievements it is apparently not the best way to teach girls. As is suggested by the MOOJ observations, this might be caused by the fact that the RME approach is hard to find in classroom practice. It turns out that this especially hits girls. Badly implemented RME means that students have to rely, in a way, on their own abilities, which has as a result that boys, given their 'natural' abilities which better fit RME, do better in this situation than girls. Maybe girls, more so than boys, are more explicitly depending on education. Applying smart strategies, acquiring estimation strategies, developing knowledge of measurements etc, must not be pursued only as goals, but schools should offer sufficient learning opportunities and enable active participation by students.

References:

Bokhove, J., Van der Schoot, F. & Eggen, Th. (1996). *Balans van het rekenonderwijs aan het einde van de basisschool 2. Uitkomsten van de tweede peiling rekenen/wiskunde in de basisonderwijs*. Arnhem: CITO.

Cobb, P., Wood. T. & Yackel, E. (1991). A constructivist approach to second grade mathematics. In E. von Glasersfeld (Ed.), *Radical Constructivism in Mathematics Education* (pp. 157-176). Dordrecht: Kluwer Academis Publishers.

Dolle-Willemsen, T.E. (1997). *Gezien onderscheid naar sekse in het basisonderwijs: Interactie als invalshoek*. Leiden: Rijksuniversiteit Leiden.

Geary, D.C. (1997). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19*, 229-284.

Hyde, J.S., Fennema, E. & Lamon, S.J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155.

Janssen, J., Van der Schoot, F., Hemker, B. & Verhelst, N. (1999). *Balans van het rekenonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 1997.* Arnhem: CITO.

Jungwirth, H. (1991). Interaction and Gender — Findings of a Microethnographical Approach to Classroom Discourse. *Educational Studies in Mathematics, 22*, 263-284.

Jungwirth, H. (1996). Symbolic Interactionism and Ethnomethodology as a Theoretical Framework for the Research on Gender and Mathematics. In G. Hanna (Ed.), *Towards Gender Equity in Mathematics Education* (pp.49-70). Dordrecht: Kluwer Academic Publishers.

Leder, G.C. (1992). Mathematics and Gender: Changing Perspectives. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 597-622). New York: Macmillan.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1997). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Shaw, L. (2002). Schools' gender gap concerns now is boys. *Seattle Times*, September 30.

Van den Heuvel-Panhuizen, M. (1997). How equally suited is realistic mathematics education for boys and girls? — A first exploration. In E. Pehkonen (ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education* (Volume 3, pp. 65-72). Helsinki/Lahti: University of Helsinki/Lahti Research and Training Center.

Van den Heuvel-Panhuizen, M., and Vermeer. H. (1999). *Verschillen tussen meisjes en jongens bij het vak rekenen-wiskunde op de basisschool* [Differences between girls and boys in mathematics at primary school]. Utrecht: CD-ß Press, Utrecht University.

Van den Heuvel-Panhuizen, M. (2001). Realistic Mathematics Education in the Netherlands. In J. Anghileri, *Principles and Practices in Arithmetic Teaching.* Innovative approaches for primary classroom (pp. 49-63). Buckingham: Open University Press.

Wijnstra, J.M. (1982). Enkele nieuwe gegevens over verschillen in toetsprestaties tussen jongens en meisjes. *Tijdschrift voor Onderwijsresearch, 7*, 43-46.

Wijnstra (Ed.), J. (1988). *Balans van het rekenonderwijs in de basisschool.* Arnhem: CITO.