12th International Congress on Mathematical Education
Program Name XX-YY-zz (pp. abcde-fghij)
8 July – 15 July, 2012, COEX, Seoul, Korea (This part is for LOC use only. Please do not change this part.)

COMPARING STUDENTS' RESULTS ON WORD PROBLEMS WITH THEIR RESULTS ON IMAGE-RICH NUMERACY PROBLEMS

Kees Hoogland, Arthur Bakker, Jaap de Koning, Koeno Gravemeijer

APS, Utrecht University, Erasmus University Rotterdam, Eindhoven University of Technology

K.Hoogland@aps.nl, A.Bakker4@uu.nl, DeKoning@ese.eur.nl,
Koeno.Gravemeijer@eseo.nl

This paper reports on an experiment comparing students' results on image-rich numeracy problems and on equivalent word problems. Given the well reported problematic nature of word problems, the hypothesis is that students score better on image-rich numeracy problems than on comparable word problems. To test the hypothesis a randomized controlled trial was conducted with 31,842 students from primary, secondary, and vocational education. The trial consisted of 21 numeracy problems in two versions: word problems and image-rich problems. The hypothesis was confirmed for the problems used in this experiment. With the insights gained we intend to improve the assessment of students' abilities in solving quantitative problems from daily life.
Numeracy, word problem, image-rich problem, randomized controlled trial, assessment.

INTRODUCTION

The experiment reported here is part of a larger research project with the aim to improve the assessment of students' mathematical literacy, in particular their abilities to solve quantitative problems from daily life. In current classroom practice, word problems are predominantly used to teach and assess these abilities. Many research findings in the past twenty years report serious difficulties in using word problems to assess these abilities (Verschaffel, Greer & De Corte, 2000; Verschaffel, Greer, Van Dooren & Mukhopadhyay, 2009). Many researchers also advocate the use of more authentic problems in such assessment (Bonotto, 2009; Frankenstein, 2009; Lave, 1992; Zevenbergen & Zevenbergen, 2009). In this paper we investigate one aspect of this broader issue: whether students can better show their abilities to solve quantitative problems in image-rich numeracy problems than in mathematically equivalent word problems. We argue that this alternative may avoid the most commonly reported difficulties that arise in the use of word problems, and may help to make numeracy problems more authentic.

THEORETICAL BACKGROUND

Numeracy and mathematical literacy

To develop students' abilities to solve quantitative problems from daily life is mentioned as a goal in almost all mathematics curricula worldwide. Kilpatrick (1996) observes "the

curriculum had shifted (…) away from an emphasis on abstract structures towards efforts to include more realistic applications, with an emphasis on the ways in which mathematics is used in daily and professional life" (p. 7). Niss (1996) speaks of "providing individuals with prerequisites which may help them to cope with life" (p. 13). If we focus on the usefulness of mathematics and its translation into education, two concepts hold a prominent place: mathematical literacy and numeracy. Over the years also other terms have been proposed to try to pinpoint the usability aspect of mathematics more precisely; examples of these terms are matheracy (D'Ambrosio, 1998), mathemacy (Skovsmose, 1998), quantitative literacy (Steen, 2001), and techno-mathematical literacies (Hoyles, Noss, Kent, & Bakker, 2010). Numeracy is used most frequently in research on adults learning mathematics (Coben, 2003) and in related international studies such as Adult Literacy and Lifeskills (ALL) Survey (Gall, Groenestijn, Manly, Schmitt, & Tout, 2003) and the Programme for the International Assessment of Adult Competencies (PIAAC). Mathematical literacy is used more often in primary and secondary education curricula (Jablonka, 2003) and in related international studies such as the Programme for International Student Assessment (PISA). What the most prominent concepts have in common is a focus on real contexts and individuals responding mathematically to problem situations. Using quantitative problems from daily life, in teaching and in assessment, is common to almost all current definitions of numeracy and mathematical literacy.

Word problems

Over the past twenty years the goals of mathematics education aiming at students' abilities to solve quantitative problems from daily life have found their way into classrooms. In common classroom practice we see a widespread use of word problems in teaching and assessing these problem solving abilities. The word problem is used as a vehicle to connect classroom practice with quantitative problems in real life. Word problems can be defined as "verbal descriptions of problem situations wherein one or more questions are raised the answer to which can be obtained by the application of mathematical operations to numerical data available in the problem statement" (Verschaffel et al., 2000, p. ix). Numerous studies have reported on students' behavior in solving word problems, for instance the superficial strategies students use, and on how they base their solutions on an association between certain salient elements of the problem situation and a certain mathematical operation (Verschaffel et al., 2000). Even more studies show that students tend not to consider possible constraints imposed by reality, and, instead, approach the word problem purely as a school mathematical problem and not as a representation of a problem from daily life (Cooper & Harries, 2003; Lave, 1992; Reusser & Stebler, 1997; Verschaffel et al., 2009). In general the classroom culture in mathematics lessons is to get to a good answer as quickly as possible and students are in many cases not expected to reflect upon their answers. This classroom culture is reinforced by word problems that are not more than poorly disguised exercises in basic operations. The idea that the explanation for this behavior by students can be found in the culture of the classroom is strongly argued by Gravemeijer (1997) on the basis of what Yackel and Cobb (1995) call the sociomathematical norms, and since then has been seen as one of the most likely explanations (Verschaffel et al., 2009). The phenomenon of superficial strategies and dissociation from reality has become known as the "suspension of

sense-making" (Schoenfeld, 1991; Verschaffel et al., 2000). This suspension of sense-making hinders students from fully showing their abilities to solve quantitative problems from daily life. And it also casts doubt on whether word problems are the right instrument for assessing these abilities. The assessment of problem-solving abilities for quantitative problems from daily life by using word problems can be summarized as troublesome.

Image-rich numeracy problems

In the present study we sought to investigate an alternative for word problems in assessing students' abilities to solve quantitative problems from daily life. With new technologies like digital cameras and on-screen presentations, it becomes easier to present quantitative problems from real life in a way that more closely resembles the real problem, without the need to simulate the complete problem situation. In this alternative approach the problem situation from reality is represented mainly with images. In this study these images are photographs. In research on alternatives for assessing mentioned abilities researchers and practitioners advocate the creation of authentic situations in the mathematics lessons to teach and assess these abilities (Bonotto, 2009; Frankenstein, 2009; Lave, 1992; Zevenbergen & Zevenbergen, 2009). Although many of the arguments for using authentic situations to teach students relevant problem-solving skills are convincing, we do not see a widespread dissemination of such practices. In many cases practical constraints in the school setting are mentioned as a major barrier. The alternative of using photographs can possibly act as a practical in-between.

The reason for using images from real life is twofold. First, from cognitive psychology and semiotics we know that depictive representations have a high inferential power, because the information can be read off more directly from the representation (Johnson-Laird & Byrne, 1991). In the solving of quantitative problems this helps students to make a relevant mental model of the situation more easily (Schnotz, Baadte, Müller, & Rasch, 2010). Second, exploratory research (Hoogland, 2007) on the skills (weak) learners show in carrying out quantitative tasks in practical vocational settings shows that students are more capable of showing their skills if the objects are at hand or if there is a close association with real problems. The hypothesis of the investigation is that using image-rich problems instead of word problems avoids most of the reported difficulties with word problems. We expect students to be less likely to fall into the trap of only looking at the verbal descriptions for clues on how to solve the "hidden" mathematical problem. We expect that, by a stronger association with the real problem, suspension of sense-making will be less common. What we focus on in this paper, however, is the conjecture that this will result in an assessment which gives a less distorted indication of the students' abilities to solve quantitative problems from daily life. In the experiment the hypothesis that students score better on image-rich numeracy problems than on equivalent word problems is tested.

METHOD

Design

The experiment was a randomized controlled trial with a 1 x 1 design. Every participant was presented with a test containing 24 items. Of these items 21 came in two versions: a word

problem version (A-version) and an image-rich version (B-version). The two versions of each item were evaluated by expert panels to establish that they are assessing the same mathematical knowledge and skills at the same level. For each participant twelve randomly chosen items were presented as a word problem and twelve were presented as an image-rich numeracy problem. The items were delivered in a random order. Through the randomization process we can assume that, for each problem, the group that answered the A-version and the group that answered the B-version have the same characteristics. By this design this should hold for the measured characteristics as well as for the characteristics that were not measured. The independent factor is the version of the problem (A or B). The dependent factor is the difference in the percentage of correct answers between the A-version and the B-version.

Participants

In the Dutch school system, primary education is for 4-12-year-olds and runs over 8 grades (K-6). In secondary education, the Netherlands has a highly streamed school system. Vmbo is a pre- vocational education stream for 12-16-year-olds, attended by around 45% of Dutch school children. Havo and vwo are the general secondary and pre-university streams that prepare children for college and university. Around 55% of Dutch children attend this stream. Mbo is the secondary vocational stream that is a follow-up to vmbo and is intended for 16-19-year-olds. In total 31,842 students from 179 schools geographically spread across The Netherlands participated in the experiment.

Table 1: number of participants

	Primary education	Secondary education			Other/ unknown	total
	Grades 5 and 6	Pre-vocational vmbo	General secondary/ pre-university havo/vwo	Secondary vocational mbo		
N	969	12,459	16,588	1,146	680	31,842

Schools and teachers voluntarily participated in this test. One of the main reasons for participating mentioned in the evaluation is the fact that the test as a whole also gave an indication of the students' level relative to the recently implemented Literacy and Numeracy Framework (Hoogland & Stelwagen, 2012). Participating schools are assumed to be representative of Dutch schools in general.

Tasks

In the experiment all participants were presented with 24 numeracy problems. Of these 24 problems, 21 came in two versions. Figures 1 provides three examples.

Figures 1 Three examples of A-versions and B-versions

For solving the problems an on-line calculator was allowed. For the total test a time limit of 60 minutes was set. All answers to the problems are numerical values. Participants typed the numerical answers they found into an empty entry field. The answers were scored by the computer.

Procedure

The test was web-based. This means the participants conducted the test on-screen at a PC with a connection to the internet. Every participant was assigned a personal activation code to start up the digital test of 24 problems. All answers delivered by the participants, including the time in milliseconds spent on the test, were recorded. After finishing the test a short digital questionnaire was delivered to each participant to collect the following additional data: gender, age, zip code (as an indicator of Social Economic Status), grade, school level, and last mark for mathematics. All data were recorded anonymously in a research database.

RESULTS

Table 2 presents an overview of the mean scores for the A-version and the B-version of the participants who were confronted with the A-version or the B-version respectively. By design this is, in both cases, about half the participants.

Table 2 Scores on all problems in two versions A and B

Number of scores of versions A	Number of scores of versions B	Mean score of versions A	Mean score of versions B	Signifi-cance	Score of versions B significantly higher than score of versions A
334600	334082	0.435	0.455	0.000	*

Note: *t* test , * significance level 0.05

Under the assumption that every participant for the A-version of a given problem i (=1, ..., 21) has the same chance of producing the right answer, the percentage of correct scores follows a binomial distribution that can be approximated by a normal distribution. Assuming this chance on producing a right answer by the participant is p_{i1}, then the mean of the sampling distribution is p_{i1} with variance $p_{i1}(1-p_{i1})/N_{i1}$, with N_{i1} indicating the number of participants answering this question. If for the B-version the same assumption holds with p_{i2} being the chance of producing the right answers, the corresponding mean and variance are p_{i2} and $p_{i2}(1-p_{i2})/N_{i2}$, with N_{i2} indicating the number of participants answering the B-version. Based on these assumptions a *t* test was conducted on the difference between the mean percentage score of A-versions and the mean percentage score of B-versions. The score on the B-version was significantly higher (Table 2).

Furthermore, for each separate problem a *t* test was conducted on the difference between the mean percentage score of the A-version and the mean percentage score of the B-version of that problem. Table 3 presents the results of these 21 tests. If neither B scores significantly better than A, nor A better than B, the difference is insignificant.

Table 3. Scores on 21 problems in two versions A and B

	Number of scores of version A	Number of scores of version B	Mean score of version A	Mean score of version B	Signifi-cance	Score of version B significantly higher than score of version A	Score of version A significantly higher than score of version B
V1	15878	15964	0.72	0.72	0.424		
V2	15986	15856	0.53	0.48	0.000		*
V3	15785	16057	0.31	0.29	0.000		*
V4	15835	16007	0.83	0.83	0.131		
V5	16038	15804	0.72	0.83	0.000	*	
V6	15775	16067	0.63	0.64	0.102		
V7	16065	15777	0.40	0.42	0.042	*	
V8	16298	15544	0.30	0.30	0.420		
V9	16069	15773	0.22	0.21	0.085		
V10	15882	15960	0.49	0.52	0.000	*	
V11	15850	15992	0.14	0.31	0.000	*	
V12	15871	15971	0.47	0.44	0.000		*
V13	15931	15911	0.62	0.64	0.000	*	
V14	15889	15953	0.04	0.05	0.080		
V15	15793	16049	0.39	0.39	0.264		
V16	15921	15921	0.80	0.82	0.005	*	
V17	15986	15856	0.80	0.79	0.016		*
V18	15847	15995	0.15	0.17	0.000	*	
V19	15932	15910	0.25	0.28	0.000	*	
V20	15925	15917	0.13	0.16	0.000	*	
V21	16044	15798	0.19	0.26	0.000	*	

Note: *t* test, * significance level 0.05

The data give a significant indication that in this test students score higher of the B-versions than of the A-versions. This was true for 10 of the 21 problems. For 4 of the 21 problems students scored significantly higher of the A-version. For 7 of the 21 problems no significant difference between the scores was found.

DISCUSSION

In the experiment the hypothesis was that students score better on image-rich numeracy problems than on equivalent word problems. The overall results in this randomized controlled trial support this hypothesis. This finding is only one step in our larger endeavor to improve the assessment of students' abilities to solve quantitative problems from daily life. We acknowledge the complexity in comparing assessment methods, especially in such a multifaceted domain as the abilities to solve problems from daily life. So we present the conclusions with great prudence.

We give two critical remarks on the way we tested the hypothesis.

1. For one third of the problems the difference in scores between the two versions was not significant and in some cases the word problem versions score significantly better. Hence, further research is needed to investigate whether the added value of using image-rich numeracy problems depends on characteristics of the problem.

2. How easy or how difficult is it to design appropriate image-rich numeracy problems? Our current follow-up research focuses on creating a typology of problems and a typology of representations of situations with which we can make better predictions of how students will score of solving quantitative numeracy problems from daily life.

Potential gains

In future research we hope to find better and more concrete explanations for the findings of this study, and why the effect appeared with some problems and not with other problems. However, the current experiment was conducted with participants most of whom are used to word problems and not used to image-rich numeracy problems. Even if using images for a given problem gives slightly worse results, it might still be useful to opt for it, because it might help students to better solve the problems they encounter in real life. Given this, the findings are promising and suggest that with better design and a better knowledge of the underlying factors that affect students' results, this approach can possibly result in a better way of assessing students' abilities to solve quantitative problems from daily life, and in that way can contribute to the justification for mathematics in education as a useful subject matter.

References

Bonotto, C. (2009). Working towards teaching realistic mathematical modelling and problem posing in Italian classrooms. In L. Verschaffel, B. Greer, W. V. Dooren & S. Mukhopadhyay (Eds.), *Words and worlds - Modelling verbal descriptions of situations* (pp.297-314). Rotterdam, Netherlands: Sense.

Coben, D. (2003). *Adult numeracy: review of research and related literature*. London, England: NRDC.

Cooper, B., & Harries, T. (2003). Children's use of realistic considerations in problem solving: some English evidence. *The Journal of Mathematical Behavior*, *22*(4)*,* 449-463.

D'Ambrosio, U. (1998). Literacy, matheracy and technoracy, the new trivium for the era of technology. In P. Gates (Ed.), *Proceedings of the first International Mathematics Education and Society Conference [MES1]* (pp. 9-11). Nottingham, England: Centre for the Study of Mathematics Education, Nottingham University.

Frankenstein, M. (2009). Developing a criticalmathematical numeracy through real real-life word problems. In L. Verschaffel, B. Greer, W. V. Dooren & S. Mukhopadhyay (Eds.), *Words and worlds - Modelling verbal descriptions of situations* (pp. 111-130). Rotterdam, Netherlands: Sense.

Gal, I., Groenestijn, M. v., Manly. M., Schmitt, M. J., & Tout, D. (2003). Numeracy in the adult literacy and lifeskills (ALL) Survey: An overview and sample items. Retrieved from http://www.ets.org/all.

Gravemeijer, K. (1997). Solving word problems: A case of modelling? *Learning and Instruction, 7*(4), 389-397.

Hoogland, K. (2007). Mind and gesture: The numeracy of a vocational student. In M. Horne & B. Marr (Eds.), *Connecting voices in adult mathematics and numeracy: practitioners, researchers and learners.* Proceedings of the 12th International Conference of Adults Learning Mathematics (ALM), Melbourne, Australia: ACU.

Hoogland, K., & Stelwagen, R. (2012). New Dutch Numeracy Framework. In T. Maguire (Ed.), *Proceedings of the 18th International Conference of Adults Learning Mathematics (ALM)*. Dublin, Ireland: ITT.

Hoyles, C., Noss, R., Kent, P., & Bakker, A. (2010). *Improving mathematics at work: The need for techno-mathematical literacies*. London, England: Routledge.

Jablonka, E. (2003). Mathematical literacy. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & F. K. S. Leung (Eds.), *Second International Handbook of Mathematics Education* (pp. 75-102). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, England: Erlbaum.

Kilpatrick, J. (1996). Introduction to section 1: Curriculum, Goals, Contents, Resources. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International Handbook of Mathematics Education*. Dordrecht, the Netherlands: Kluwer Academic Publishers

Lave, J. (1992). Word Problems: A Microcosm of Theories of Learning. In P. Light & G. Butterworth (Eds.), *Context and cognition: Ways of learning and knowing* (pp. 74-92). New York, NY: Harvester Wheatsheaf.

Niss, M. (1996). Goals of mathematics teaching. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International Handbook of Mathematics Education.* Dordrecht, Netherlands: Kluwer Academic Publishers

Reusser, K., & Stebler, R. (1997). Every word problem has a solution – The social rationality of mathematical modeling in schools. *Learning and Instruction*, *7*(4), 309-327.

Schnotz, W., Baadte, C., Müller, A., & Rasch, R. (2010). Creative thinking and problem solving with depictive and descriptive representations. In L. Verschaffel, E. d. Corte, T. d. Jong & J. Elen (Eds.), *Use of representations in reasoning and problem solving - Analysis and improvement* (pp. 11-35). London, England: Routledge.

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In J. F. Voss, D. N. Perkins & J. W. Segal. (Eds.), *Informal reasoning and education*. Hillsdale, NJ: Lawrence Erlbaum.

Skovsmose, O. (1998). Linking mathematics education and democracy: Citizenship, mathematical archaeology, mathemacy and deliberative interaction. *Zentralblatt für Didaktik der Mathematik*, *98*(6), 195-203.

Steen, L. A. (Ed.). (2001). *Mathematics and democracy, The case for quantitative literacy*. Washington, DC: NCED, The Woodrow Wilson National Fellowship Foundation.

Verschaffel, L., Greer, B., Van Dooren, W. & Mukhopadhyay, S. (Eds.). (2009). *Words and worlds - Modelling verbal descriptions of situations.* Rotterdam, The Netherlands: Sense.

Verschaffel, L., Greer, B., & De Corte, E. (Eds.). (2000). *Making sense of word problems.* Lisse, The Netherlands: Swets & Zeitlinger Publishers.

Yackel, E., & Cobb, P. (1995). Classroom sociomathematical norms and intellectual autonomy. In L. Meira & D. Carraher (Eds.), *Proceedings of the Nineteenth International Conference for the Psychology of Mathematics Education* (Vol. 3, pp. 264-271). Recife, Brazil: Program Committee of the 19th PME conference.

Zevenbergen, R., & Zevenbergen, K. (2009). The numeracies of boatbuilding: New numeracies shaped by workplace technologies. *International Journal of Science and Mathematics Education, 7*(1), 183-206.