PSYCHOMETRIKA—VOL. 74, NO. 2, 367–374
JUNE 2009
DOI: 10.1007/S11336-009-9124-1

HOW TO MEASURE AND EXPLAIN ACHIEVEMENT CHANGE IN LARGE-SCALE ASSESSMENTS: A REJOINDER

MARIAN HICKENDORFF, WILLEM J. HEISER, AND CORNELIS M. VAN PUTTEN

LEIDEN UNIVERSITY

NORMAN D. VERHELST

CITO, NATIONAL INSTITUTE FOR EDUCATIONAL MEASUREMENT

In this rejoinder, we discuss substantive and methodological validity issues of large-scale assessments of trends in student achievement, commenting on the discussion paper by Van den Heuvel-Panhuizen, Robitzsch, Treffers, and Köller (2009). We focus on methodological challenges in deciding what to measure, how to measure it, and how to foster stability. Next, we discuss what to do with trends that are found. Finally, we reflect on how the research findings were received.

Key words: mathematics education, assessment of trends, complex division, curriculum assessment.

1. Introduction

The most important aim of large-scale educational assessments (international, such as TIMMS and PISA, and national, such as NAEP in the US and PPON in the Netherlands) is to report on the output of the educational system. Two aspects are particularly of interest. The first aspect is a description of students' learning outcomes: what do students know, what problems can they solve, to what extent are educational standards reached, and to what extent are there differences between subgroups (such as different countries, or boys and girls within a country)? The second aspect concerns trends: to what extent are there changes in achievement level over time?

Assessments are descriptive by nature: explanations for differences found would usually require further study (Meelissen & Drent, 2008). In our research (Hickendorff, Heiser, Van Putten, & Verhelst, 2009), we undertook such a further study, focusing on a specific trend found in Dutch mathematics assessments: achievement on complex division problems showed a large decrease between 1987 and 2004 (Janssen, Van der Schoot, & Hemker, 2005). We tried to contribute to an explanation of this trend by relating (changes in) solution strategies to achievement results of the two most recent assessments of 1997 and 2004.

Van den Heuvel-Panhuizen, Robitzsch, Treffers, and Köller (2009) discussed substantive and methodological validity issues of large-scale assessments of change in students' achievement, and raised several serious concerns about studies like ours. In this rejoinder, we will reflect upon three themes: methodological challenges for large-scale educational assessments (what to measure, how to measure it, and how to foster stability), what to do with trends found, and how the research findings were received.

Requests for reprints should be sent to Marian Hickendorff, Division of Methodology and Psychometrics, Leiden University Institute for Psychological Research, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: hickendorff@fsw.leidenuniv.nl

2. Educational Assessments: Methodological Challenges

Large-scale assessments are a comprehensive and difficult enterprise. There are all sorts of methodological challenges to obtaining reliable, valid, and comparable measurements of educational outcomes over time, given a changing educational environment. Generally speaking, careful planning and documentation and using standardized procedures are essential. Two methodological considerations are central in setting up an assessment. The first is what to measure and how to measure it: the learning outcomes tested should cover the curriculum content (what) in an appropriate format (how). Second, utmost care should be taken to foster stability in successive assessments. Particularly important for reliable trend assessment is to make as few changes as possible between consecutive assessments. In the following, we will discuss these two issues in more detail, with the Dutch national assessments and our study as an illustration.

2.1. What to Measure and How to Measure It?

Educational assessments usually measure learning output of the curriculum. Most scholars distinguish between the intended and the enacted curriculum (for example, Porter, 2006; Stein, Remillard, & Smith, 2007). The *intended* curriculum is based on written documents such as educational standards and textbooks, while the *enacted* curriculum is the actual instruction taking place in the classroom. The learning outcomes to be measured should be aligned with both curricula (Porter, 2006): it should cover content that students should have learned (the intended curriculum) and had an opportunity to learn (the enacted curriculum).

In designing such a learning outcomes test, one has to start with a framework of the content area at stake. For example, the framework of TIMMS (Trends in International Mathematics and Science Studies) is based on the broadly defined (international) school curriculum of what students should have learned (Mullis, Martin, & Foy, 2008), while PISA (Programme for International Student Assessment) covers content that goes beyond the school curriculum, by assessing the functional use of acquired mathematics knowledge and skills in realistic situations (OECD, 2004). The Dutch national assessments (PPONs) aim to evaluate the educational system and, therefore, base their mathematics framework on indicators of the intended and enacted mathematics curriculum: the governmental educational standards for primary education, mathematics textbooks used, and didactical considerations (Van der Schoot, 2008).

When operationalizing the curriculum content into test items, two aspects are relevant: the topic and the cognitive demand (Porter, 2006). It is very important to define both aspects carefully, which we will illustrate by taking what was measured in our study (Hickendorff et al., 2009) as an example. In this study, the topic at stake was *complex division*, defined as the ability of students to solve division problems with large numbers in their own way, in which the use of scrap paper is allowed. Van den Heuvel-Panhuizen et al. (2009), however, used a different label (written division) as well as a different definition of this topic: they restricted it to "the ability to carry out a written calculation" throughout their discussion. This curtailment is not in agreement with the intentions of the PPON-researchers, who state that [translated from Dutch] "*Students are free in their choice of a solution strategy*." (Van der Schoot, 2008, p. 21). The fact that Van den Heuvel-Panhuizen et al. assumed a different (incomplete) definition of the content at stake caused several of their theoretical and methodological concerns raised, as will become clear from the following.

A first concern raised was that many of the problems presented are not typically solved by written arithmetic, nowadays (Van den Heuvel-Panhuizen et al., 2009). We beg to differ on this speculative statement in the first place, since it does not seem to be warranted by empirical data. In the second place, however, we argue that, even if this would be true, it would not be a problem for the construct validity. That is, problems with easier numbers that may be solved by mental calculation do fit in our broader defined topic. Moreover, including a set of problems ranging

in cognitive demands or difficulty level ensures broad coverage of a topic, such as, for example, also argued in NAEP, the US National Assessment of Educational Progress (National Assessment Governing Board, 2006). In fact, the issue of mental calculation is explicitly addressed by the PPON-researchers [translated from Dutch]: "*There are problems included that can clearly be solved mentally. However, one would expect students to seize the opportunity to make use of scrap paper to increase their chances of finding the correct solution*." (Janssen et al., 2005, p. 236).

A similar argument can be made regarding another concern raised by Van den Heuvel-Panhuizen et al. (2009) on the representativeness of the problems for the construct (as they defined it). It is beyond discussion that an item set intended to measure some aspect of learning output should take into account as many relevant item features as possible, to ensure adequate content validity. Practical constraints limit the number of items per topic, though. Distributing item features in a balanced way to allow testing for independent effects of distinct item features is not feasible in large-scale assessments. Not only would it require many more items per topic, it is also beyond the scope of reporting on educational outcomes. However, assessment results could yield hypotheses on effects of item features that can be tested in further experimental studies, such as on the differences between bare number and context problems.

So, *what* to measure is a matter of covering the content of the intended (educational standards) and/or the enacted curriculum (opportunity to learn), with respect to topics as well as cognitive demands. The next important methodological consideration is *how* to measure it: the format of assessment. Without doubt, choice of format affects results. In our view, format and instruction should suit varying classroom circumstances in the best possible way, given practical considerations. Usually, achievement is measured with classroom paper-and-pencil tests, while information on students' approaches to problems can be obtained by think-aloud protocols with a smaller number of students (Porter, 2006).

These two types of assessment formats were included in the Dutch assessment as well (Janssen et al., 2005). Not surprisingly, they yielded different results regarding strategies used and performance (Van Putten & Hickendorff, 2006). These differences led Van den Heuvel-Panhuizen et al. (2009) to question the validity of the classroom test for measuring performance, arguing that the prompt for written strategies was not sufficient. Although we showed before that the classroom test was not intended to restrictively measure written strategies, we also argue that attributing the differences found only to the prompt for written strategies presents a misleading picture. There are many more differences between classroom testing for achievement and individual interviews meant to get insight into solution strategies. For one, the interviewers were instructed that they could help students.

These and other findings led us to experimentally test the effect of prompting for written strategies. We partially implemented the Choice/No-Choice methodology, in a classroom administration of a paper-and-pencil test (Hickendorff, Van Putten, Verhelst, & Heiser, 2009). We found that the probability of a correct answer was raised with an average of 16 percentage points by forcing to write down the solution strategy, but *only* for those students who used a mental strategy when they were free to choose. Therefore, when all students were considered, a much smaller performance difference was found between free strategy choice and forced written strategies. We think that these results give better insight in the effect of prompting to use a written strategy, than comparing paper-and-pencil tests with individual interviews. Moreover, we argue that free strategy choice much better suits current classroom practice and, therefore, also the educational curriculum in the Netherlands, since a diverse strategy repertoire and flexible use of strategies are central points of realistic mathematics education (Van den Heuvel-Panhuizen et al., 2009). Forcing the use of written strategies would yield measurements that would not reflect how students learn mathematics nowadays, which would be problematic for validly assessing educational outcomes (Van Putten, 2008).

2.2. *How to Foster Stability?*

Since reliable trend measurement is one of the major goals of (inter)national assessments, fostering stability between consecutive assessments is one of the leading considerations in designing an assessment. Essentially, all possible factors that may affect results should be as similar as possible, to rule out these sources of uncertainty. These factors are, among others, the sample of students, time of measurement (especially with respect to educational practice in schools), and the assessment format and instruction. Changing any of these factors would require (expensive) bridging studies (Mazzeo & von Davier, 2008). As a result, choices made in the design of a first assessment affect the design of all consecutive assessments.

Given the changing educational environment, the CITO national assessments have tried to keep as many factors constant as possible. For example, in comparing the 1997 with the 2004 results, both assessments included a large representative sample of students, measured at the same time point in their elementary school trajectory (at the end of their grade 6 school year) with the same type of problems and instructions. The only difference was the change from test booklets containing items on one domain of mathematics, to booklets with items on more than one mathematical domain.

Unfortunately, Van den Heuvel-Panhuizen et al. (2009) make a comparison in which these requirements of keeping relevant factors constant are not met, when comparing achievement results of the end of the school year in 1997 with results halfway the school year in 2004. There are many sources of incomparability, mainly because the educational context of these two time points in grade 6 is quite different. In the Netherlands, the large majority of schools let their students take the CITO end of primary school test for an objective assessment of the students' scholastic achievement. This test always takes place halfway grade 6 (in February). A very likely (and often reported) side-effect is that in the first half of grade 6, schools spend a lot of time to the areas being tested, including mathematics. In the second half of grade 6, schools may spend relatively more time on other areas, such as culture and creativity. This may have caused the 2004 finding that students mid-grade 6 performed better than at the end of grade 6. This pattern was found for almost all aspects of mathematics being measured (Janssen et al., 2005), an effect comparable to the "summer drop," the consistent finding that achievement test scores decline over summer holidays.

Of course, this phenomenon merits a discussion on school practice in grade 6 (in fact, currently being held by the Dutch association for primary education [PO-raad in Dutch]). However, the pattern does make clear that a comparison between end grade 6 results of 1997 and mid-grade 6 results of 2004 is confounded by educational context effects. Moreover, it is much more likely that in 1997 the same homogeneous achievement decrease between halfway and the end of the school year was present, instead of Van den Heuvel-Panhuizen et al. (2009, Section 3.2) far-fetched reasoning that the 2004 cross section was a more unfortunate snapshot of individual performance trajectories than the 1997 cross section was.

This brings us to the issue of individual performance trajectories. Studying these would yield informative additional data. However, the fact that this would require repeated testing of students comes with too many practical problems for it to be part of large-scale assessments. Perhaps an alternative way to study individual trajectories would be by means of CITO's monitoring and evaluation system, in which students are tested each half year, starting in first grade. It might be a promising topic for future cooperation between didactic experts, educational researchers, and psychometricians, focusing more on processes of individual change and diagnostic issues than on cross-sectional measurements of changes in the output of the educational system.

We have some final remarks on the linking issues raised by Van den Heuvel-Panhuizen et al. (2009) regarding the PPONs of 1997 and 2004. The first issue concerns the number and appropriateness of the anchor or link items. Of course, the more anchor items are included, the more stable results can be obtained. Although the complex division achievement scale was based

on only four anchor items, the total performance scale as published in Janssen et al. (2005) included both complex division as well as complex multiplication problems, with a total number of 9 link items. Therefore, we feel quite confident that this link is sufficient. Moreover, the four anchor items on division each showed a similar achievement decrease of 15 to 20 percentage points between 1997 and 2004 (Hickendorff et al., 2009, Table 1). The second issue concerns item drift or DIF, an important and relevant potential threat. When the time span gets larger, as in comparing results of 2004 with 1987, the chance of unstable items increases, making these item less suited or unsuited for comparisons (Janssen et al., 2005). Therefore, one should always test for DIF, and one should also be cautious in making comparisons covering a large time span. In the common measurement scale for the division problems of 1997 and 2004, DIF with regard to year of assessment was not present in the four anchor items.

3. We Found a Trend: What Next?

Assessing that the achievement level has changed between time points is one thing, trying to explain such a trend is another thing. The latter problem implies finding and including relevant variables that can influence performance and performance change. These variables can be student characteristics such as gender, SES, migration background, intelligence, motivational variables, and other psychological variables. Furthermore, school and curriculum variables may affect scholastic achievement: for example, the relative importance of different subdomains in the intended and enacted curriculum, teacher variables, and availability of remedial teaching.

One very important variable discussed by Van den Heuvel-Panhuizen et al. (2009) is opportunity-to-learn (OTL). In fact, Hiebert and Grouws (2007) argue that OTL is the single most important predictor of student achievement. OTL depends not only on the curriculum materials (the intended curriculum), but also to a large extent on teachers' interpretation of these materials and interactive influences of students and teachers (Stein et al., 2007). Incorporating indicators of OTL of different (RME-based) textbooks as suggested by Van den Heuvel-Panhuizen et al., as well as OTL-indicators based on actual classroom instruction practices, may yield an important contribution to trend explanation. These types of indicators did not receive much attention in the assessment reports. Regarding mathematics textbook, many schools changed textbooks between 1997 and 2004 due to the introduction of the new monetary unit (the Euro) in 2002. As a result, it was questioned to what extent the textbook used in Grade 6 in 2004 reflected the full instruction these students had experienced in their school trajectory.

An important variable reflecting the enacted curriculum came from a teacher survey in the 2004 assessment, in which teachers were asked for each operation (addition, subtraction, multiplication, and division) whether they predominantly instructed the traditional algorithm, realistic strategies, or both. Since this information was not collected in 1997, it could not be incorporated as a covariate in our IRT model to explain achievement change between 1997 and 2004. For 2004, though, it can be related to strategy use. The main strategy instructed turned out to be a strong predictor of the main strategy used, in particular with respect to the traditional algorithm: students not instructed the traditional strategy almost never used it. This is an example of the effect of opportunity-to-learn or the enacted curriculum on student learning outcomes. Due to this strong relation, we expect that, with the actual strategy used incorporated as explanatory variable in the model, including teacher instruction would not result in an additional contribution explaining achievement change. It also exemplifies how the enacted curriculum can differ from the intended curriculum, since none of the textbooks used in this study includes the traditional algorithm for long division.

Moreover, we argue that studying changes in solution strategies and incorporating these as explanatory variables in the IRT models, yields very valuable information. It would be very

informative to study solution strategies in other domains of mathematics as well and relate them to trends that were found. However, it is worth noting that for the topic of complex division, even after accounting for shifts in strategy use between 1997 and 2004, still a substantial performance decrease (within each strategy) was left that could not be explained by the currently available variables. Therefore, it is necessary that future studies try to find other variables, such as other indicators of the enacted curriculum, that may contribute to explaining the changes found.

4. How Were the Findings Received?

National assessment results showed a change in competence profile of Dutch primary school students in the period 1987–2004 (Janssen et al., 2005). Van den Heuvel-Panhuizen et al. (2009) argue that this change in competence profile is in line with the proposed reform of mathematics education. In that respect, the achievement decline in complex operations may be just a trade-off of more attention (and educational time) devoted to other aspects of mathematics, so it would be nothing to worry about. However, we think this conclusion does not do justice to the fact that the educational standards for complex arithmetic, set by the Dutch government, are far from being reached (Van der Schoot, 2008). In addition, a recent expert panel report on what students should have learned at the end of primary school (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008) states that solving complex arithmetic problems is an important part of mathematics education. Therefore, we argue that smoothing over the reported problems in this particular aspect of school mathematics does not contribute to meeting societal demands of mathematics education.

Specifically, when results on complex division are considered, Van den Heuvel-Panhuizen et al. (2009) are contented with the result reported that the so-called realistic (chunking-based) strategies for solving complex division problems were found to be as accurate as the traditional algorithm. Although this result holds for weak and strong mathematical achievers but not for the average achievers (Hickendorff et al., 2009, Figure 5), it is a promising result for the reform-based approach to complex division. However, it presents only half of the picture. The other half is as follows: the traditional algorithm did not seem to be replaced by the realistic strategies, but instead by the far less accurate strategy of answering without any written working. Moreover, each strategy on its own (including the realistic strategies) had become less accurate over time. We believe that these two additional findings cannot be ignored when qualifying the result that a realistic strategy was found to be as successful as the traditional strategy as something to be contented with. Instead, we believe that the complete picture of results is far from satisfying from an educational perspective.

In addition, in response to the increase in strategies without written workings (mental calculation, most likely) instead of an increase of realistic strategies, Van den Heuvel-Panhuizen et al. (2009, Section 2.1) argue that "*with increased number sense, mental calculation and estimation, one no longer needs the written chunking strategy*." However, they seem to overlook our empirical finding that the (supposed) mental calculation strategy is far less accurate than any of the written strategies, especially for the weak achievers in mathematics. Moreover, national assessment results show that, contrary to what Van den Heuvel-Panhuizen et al. claim several times, achievement on (forced) mental multiplication and division did *not* increase in the period 1987–2004 (Janssen et al., 2005; for the period 1997–2004 also visible in Figure 3 in Van den Heuvel-Panhuizen et al.). So, we argue that evaluating this strategy shift to more mental calculation as positive is not justified by these empirical findings on achievement. In our view, credibility is at stake if we are just pleased that students shift their strategy use, irrespective of the question whether they are still able to solve the problems.

5. Concluding Remarks

Large-scale educational assessments, like any psychometric measurement, will always be surrounded by uncertainty and practical constraints. As a consequence, assessments should be designed and carried out very carefully, and choices and decisions that were made and how these may affect the validity, reliability, or generalizability of the results should be discussed explicitly. In our view, the national assessments of primary education in the Netherlands as carried out by CITO are unique in their kind: there are few other countries that have such a broad, long-term, large-scale assessment enterprise, trying to keep as many confounding variables under control given practical considerations. Of course, we do not mean to imply that there are no limitations. For one, as in all assessments, the results are necessarily restricted to what was measured and how it was measured. Therefore, it is good practice that reports are very clear on what was exactly done and why, so that everyone can evaluate the findings and its implications for him or herself. Another limitation is that assessments are surveys and therefore descriptive by nature. As a result, only correlational associations of explanatory variables with student learning outcomes can be established. Such relations may provide starting points for further experimental studies, in which hypotheses can be tested explicitly. Among other things, we suggest performing studies into effects of item format, item characteristics, and test instruction, preferably in an international comparative setting. Furthermore, it would be very important to try studying the entire chain of curricular materials, teacher interpretation, curricular enactment, and student learning (Stein et al., 2007).

However, even given their limitations, these national assessment results are valuable and should be taken seriously by teachers and teacher's educators, by the government, and by mathematics educators and researchers. We argue that Van den Heuvel-Panhuizen et al. (2009) have been very keen in listing all sorts of potential threats to the validity of the findings. However, we showed that the concerns raised on the construct validity are based on an incorrect restriction of the definition of the topic at stake, while the concerns about assessment factors were mostly based on unsuitable comparisons and, therefore, unlikely to have affected trend results. Moreover, we think that such an open-minded and observant point of view is lacking in their reception of the outcomes of the assessments as well as in their discussion of the implications for mathematics education.

We cannot escape the conclusion that something is going on in Dutch mathematics achievement. Many things go well: national results of PPON showed improvements on some aspects of mathematics, and the international comparative studies TIMMS (Meelissen & Drent, 2008; Mullis et al., 2008) and PISA (De Knecht-Van Eekelen, Gille, & Van Rijn, 2007) consistently showed that Dutch students perform at the top level internationally. On the downside, however, TIMMS 2007 and PISA 2006 each reported a significant negative trend over time for Dutch mathematics performance. In addition, national assessments consistently reported that on many aspects of mathematics the educational standards were not reached, and also that on some aspects performance decreased considerably. Our study gives some important insights into one aspect that is not going well: solving complex division problems for which the use of scrap paper is allowed. It showed that the decline in achievement is related to a shift in strategy use as well as to an autonomous performance decline within each strategy. Although these findings can only be the starting point of a comprehensive evaluation of assessment results, they are important for those involved in mathematics education. Further research into what caused this and other negative trends is crucial for improving mathematics education.

References

De Knecht-Van Eekelen, A., Gille, E., & Van Rijn, P. (2007). *Resultaten Pisa-2006. Praktische kennis en vaardigheden van 15-jarigen [Pisa 2006 results. Functional knowledge and skills of 15-year-olds]*. Arnhem: CITO.

Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008). *Over de drempels met rekenen. Consolideren, onderhouden, gebruiken en verdiepen [Crossing the thresholds with mathematics. Strengthen, maintain, use, and deepen]*. Enschede: Expertgroep Doorlopende Leerlijnen Taal en Rekenen.

Hickendorff, M., Heiser, W.J., Van Putten, C.M., & Verhelst, N.D. (2009). Solution strategies and achievement in Dutch complex arithmetic: Latent variable modeling of change. *Psychometrika, 74*(2). doi:10.1007/s11336-008-9074-z.

Hickendorff, M., Van Putten, C.M., Verhelst, N.D., & Heiser, W.J. (2009). Individual differences in strategy use on division problems: mental versus written computation. *Manuscript submitted for publication*.

Hiebert, J., & Grouws, D. (2007). The effects of classroom mathematics teaching on students' learning. In F.K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte: Information Age Publishing.

Janssen, J., Van der Schoot, F., & Hemker, B. (2005). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4 [Fourth assessment of mathematics education at the end of primary school]*. Arnhem: CITO.

Mazzeo, J., & von Davier, M. (2008). *Review of the Programme for International Student Assessment (PISA) test design: Recommendations for fostering stability in assessment results*. Paris: OECD Education Working Papers (EDU/PISA/GB(2008)28).

Meelissen, M.R.M., & Drent, M. (2008). *TIMMS-2007 Nederland. Trends in leerprestaties in exacte vakken van het basisonderwijs [TIMMS 2007 the Netherlands. Trends in achievement in mathematics and science in primary education]*. Enschede: Twente University.

Mullis, I.V.S., Martin, M.O., & Foy, P. (2008). *TIMMS 2007 international mathematics report. Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Boston: Boston College TIMMS & PIRLS International Study Center.

National Assessment Governing Board (2006). *Mathematics framework for the 2007 National Assessment of Educational Progress*. Washington: US Department of Education.

OECD (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.

Porter, A.C. (2006). Curriculum assessment. In J.L. Green, G. Camilli, & P.B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 141–160). Mahwah: Lawrence Erlbaum Associates.

Stein, M.K., Remillard, J., & Smith, M.S. (2007). How curriculum influences student learning. In F.K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319–370). Charlotte: Information Age Publishing.

Van den Heuvel-Panhuizen, M., Robitzsch, A., Treffers, A., & Köller, O. (2009). Large-scale assessments of change in student achievement: Dutch primary school students' results on written division in 1997 and 2004 as an example. *Psychometrika, 74*(2). doi:10.1007/s11336-009-9110-7.

Van der Schoot, F. (2008). *Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPON [A summary overview of 20 years of national assessments of the level of education]*. Arnhem: CITO.

Van Putten, C.M. (2008). De onmiskenbare daling van het prestatiepeil bij de bewerkingen sinds 1987—een reactie [The unmistakable decline of the complex arithmetic achievement level since 1987: A reaction]. *Reken-wiskundeonderwijs: onderzoek, ontwikkeling, praktijk*, *27*(1), 35–40.

Van Putten, C.M., & Hickendorff, M. (2006). Strategieën van leerlingen bij het beantwoorden van deelopgaven in de periodieke peilingen aan het eind van de basisschool van 2004 en 1997 [Students' strategies when solving division problems in the PPON test end of primary school 2004 and 1997]. *Reken-wiskundeonderwijs: onderzoek, ontwikkeling, praktijk*, *25*(2), 16–25.

Manuscript Received: 13 FEB 2009
Final Version Received: 12 MAR 2009
Published Online Date: 22 APR 2009