Measurement estimation in primary school

Ruwisch, Silke; Heid, Marleen; Weiher, Dana Farina

Published in:
Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

MEASUREMENT ESTIMATION IN PRIMARY SCHOOL: WHICH ANSWER IS ADEQUATE?

Silke Ruwisch, Marleen Heid, Dana Farina Weiher

Leuphana University Lueneburg, Germany

Measurement estimation is seen as an important part of mathematics learning, although still very little is known about children's abilities in this respect. To make matters worse, criteria for the adequacy of estimates are arbitrarily chosen and differ in studies on this topic. If teachers have to evaluate students' estimation performances, they need criteria, too. In this paper, we first present some of those studies and their criteria for adequacy. These criteria are evaluated and prepared for discussion by applying them to data from an interview study with 4th grade students estimating length and capacity. This empirical basis for discussion is complemented by data of expert opinions about the items and children's results.

INTRODUCTION

Mathematics in primary school is usually seen as a discipline of precision. Children have to learn how to calculate correctly. Even lessons in measurement encourage students to measure as accurately as they can. For this reason, students gain a one-sided picture of mathematics as a discipline. But as Freudenthal stated already in 1978, there are two different 'worlds of mathematics' that have to be known by students: one in which precision is virtuous and one in which it is vicious.

Depending on the context and the questions one tries to answer, numbers and measures have to be more or less accurate. Whereas calculation is a procedure in the exact world of mathematics, estimation is an integral part of the second view of mathematics. Therefore, estimation recently gained more attention in the curricula of different countries such as Germany or Taiwan (see Huang, 2014).

In our study, we are mainly interested in strategies fourth-graders use to estimate length and capacity (see Ruwisch & Heid, 2015). Interpreting the answers quantitatively as well, we realised that there is no clear criterion that allows us to decide, whether an estimation is a good one. Although some studies about measurement estimation had also included a quantitative analysis (e.g. Swan & Jones, 1980; Hildreth, 1980; Siegel, Goldsmith, & Madson 1982; Clayton 1992; Jones, Forrester, Gardner, Andre, & Taylor, 2012; Huang, 2014), the authors used different criteria for this decision. This fact motivated us to take a closer look at those criteria and evaluate them here by applying them to our data.

THEORETICAL BACKGROUND

Estimation processes in mathematics lessons can be divided into three different contents: computational estimation, numerical estimation, and measurement estimation (O'Daffer, 1979; Sowder, 1992). We will restrict ourselves to the last one

2015. In Beswick, K.., Muir, T., & Wells, J. (Eds.). *Proceedings of 39th Psychology of Mathematics Education conference*, Vol. 4, pp. 113-120. Hobart, Australia: PME.

in this paper. Measurement estimation is a mental process that is thought to be analogous to real measurement processes but without handling a measurement tool (Bright, 1976; Sowder, 1992). Most research in measurement estimation is focused on lengths (see Sowder, 1992; Jones et al., 2012). Since our own study also deals with length and capacity, we will mainly focus on items in these measurement areas.

Adequacy of estimated measures: the terminology

In the literature dealing with the adequacy of estimated measures, there is also no agreement concerning the terminology. Most researchers use 'accuracy' (e.g., Swan & Jones, 1980; Siegel et al., 1982; Jones et al., 2012; Huang, 2014). In our opinion, this term overemphasises the aspect of precision and correctness. Huang (2014) also uses 'acceptability', a term that already includes the scope for decision making by the researcher. Other researchers use 'reasonable estimates' (Clayton, 1992) or 'reasonableness' (Siegel et al., 1982) as well, but even these terms differ in their meaning. Whereas Siegel et al. (1992) call comprehensible estimations 'reasonable', Clayton (1992) emphasises the complex situation that has to be taken into account when deciding the adequacy of estimations. The term 'adequacy' which is used in this paper, focuses on the equivalence between the estimation and the real measure, and may also evoke the association of precision. In German the word '*angemessen*' is used as a synonym for 'adequate'. '*Angemessen*' literally means 'to be measured with reference to something else'. In this sense, the adequacy of estimations is dependent on a reference point. So one of our questions is: Which reference point(s) can be useful to decide, whether an estimated measure is adequate?

Criteria for adequacy of estimated measures in the literature

In 1980, Swan and Jones reported about their measurement estimation studies from the seventies. 780 elementary school children (Grades 4 to 6) participated in 1971, and 304 did so in 1977. Every child had to provide written answers to eight estimation problems. Four of these problems dealt with length: "two distance intervals one of which was between 50 and 75 meters in length, the other 5 to 10 meters in length. […] two heights, one of which was about 20 meters tall, and the other shorter (such as a flagpole)." (Swan & Jones, 1980: 299). As the authors admitted, they arbitrarily judged an estimate within a maximum deviation of 25 % from the real value as 'accurate'. Although the students performed better in 1977, only 13 to 39 % gave an 'accurate' estimate of the lengths under these conditions. Junior high school students (Grades 7 to 8) performed significantly better but still poor: 21 to 50 %. Since the authors did not present their raw data, no conclusions about the deviations from the real values can be drawn.

In 1980, Hildreth published his PhD dissertation about the use of estimation strategies for length and area. Since we were not able to access the entire dissertation, raw data and detailed results of this study with 24 fifth-graders, 24 seventh-graders, and 24 college students cannot be reported here. Nevertheless, it can be stated that Hildreth measured the estimation ability by "the number of items on which the relative error was less than 1/3" (phdtree.org/pdf/24304583). Thus, a good estimation deviates within a 33 % range from the real value.

In 1982, Siegel et al. reported about skills in estimating length and numerosity. Six different types of estimation problems in four contexts were presented to 20 children of each grade (Grades 2 to 8). Two problem types dealt with numbers only, two others with length only. The remaining two problem types asked for a combination of estimating numbers as well as lengths and to calculate them. Siegel et al. differentiated between 'accuracy' and 'reasonableness'. Whereas an 'accurate' estimation was defined as a maximum deviation of 50 % from the actual value, the authors scored an estimation 'reasonable', if it was "plus or minus an order of magnitude of the actual value" (217). Since the authors were interested in the different problem types no overall data were given in the paper. Unreasonable answers only were given if the estimation process got difficult (e.g., in the combined estimation problem type). Nevertheless, benchmark problems dealing only with length were performed much better than the other problem types – no unreasonable answers were observed here – and older students performed better in all problem types than did younger students. Again, no raw data are given, so no conclusions about the adequacy of the criteria are possible.

In a recent study Huang (2014) used a two-step process to score the estimated measures of 72 fourth-, fifth-, and sixth-graders. In her study she presented 12 problems that required the estimation of length and area. In scoring the children's answers, she differentiated between 'accurate' and 'acceptable'. 'Accuracy' was defined as a maximum deviation of 10 % from the real value and scored by 2 points, whereas 'acceptability' was defined as a maximum deviation of 25 % from the real value and scored by 1 point. In length-estimation the children could achieve a maximum of 12 points. The results show that on average fourth-graders achieve of 5.91 points, whereas fifth- and sixth-grader did slightly but not significantly better. Again, no conclusion about the adequacy of the evaluating process is possible due to the fact of missing raw data. Nearly the same process is used by Hogan and Brezinski (2003). They decided to use a three-step scoring: 3 points for an answer within a range of 10 %, 2 points within 10 to 20 % and 1 point within 20 to 30 %. Since measurement estimation was a very small part of the whole study with college students as participants, no further information will be presented here.

Although there are some other suggestions how to evaluate the adequacy of estimations – Lörcher (2000) defined accuracy by an interval from the half to the double of the real value; Clayton (1992) proposed a logarithmic model, but applied it to numerosity only – we will focus on the criteria mentioned above.

METHOD

Measurement estimation tasks

Our tasks for estimating length and capacity were constructed with reference to Bright's (1976) typology of requests in estimating length. First of all, it can be differentiated if a suitable measure has to be given to a representative or if a suitable representative has to be found to a given measure. In each case the (possible) representatives can be physically present or absent as well as the unit itself may be visible or not (for more details see Ruwisch & Heid, 2015).

If the representatives are given and physically present, it can clearly be said how long, wide, tall or high they actually are, when a subject is asked to estimate their lengths. This applies equally for the estimation of the capacity of objects: If the representatives are given and physically present, it can clearly be said how much capacity they actually take.

Therefore, the answers to these tasks will be chosen for discussing our question concerning the adequacy of estimations given by the children. The following objects were presented to estimate their lengths: the diameter of the head of a wooden bug (5 mm), the length of a piece of chalk (8 cm), the length of a book with an unusual format (46 cm), the height of the table (70 cm), and the height of the room (3 m). The following objects were presented to estimate their capacities: a test tube (10 ml), a small glass (100 ml), a vase (300 ml), a carafe (500 ml or 1 l), and a big pot (3.5 l).

Sample

One hundred and thirty fourth-graders from 13 primary schools in the north of Germany were involved in this part of the study, but not every child estimated all tasks given above. As the data in Table 1 show, the total numbers of answers differ from 77 (test tube) to 128 (table).

Tasks for estimating length			Tasks for estimating capacity		
Item	Length	Number of answers	Item	Length	Number of answers
bug	5 mm	117	test tube	10 ml	77
chalk	8 cm	112	glass	100 ml	116
book	46 cm	95	vase	300 ml	115
table	70 cm	128	carafe a)	500 ml	80
room	3 m	88	carafe b)	1 l	44
			pot	3.5 l	117

Table 1: Total numbers of answers to each item

Although estimation should be part of the curriculum since 2004, none of the teachers participating in this study fostered it in their classes. All students were familiar with the measurement of length, and had already gone through one unit about capacity during this school year. All children were interviewed individually during the second half of the school-year; the whole interviews lasted about 20 to 25 minutes (see Ruwisch & Heid, 2015 for more details).

On the purpose of comparison, 17 mathematics educators who participated in a conference workshop estimated themselves the ten items given above. Afterwards, they were asked to evaluate given ranges of deviations. They should choose that range they think to be adequate for the evaluation of estimates given by 4th grade students.

Data

Table 2 shows the minimum and maximum estimations that were given by any child.

Tasks for estimating length				Tasks for estimating capacity			
Item	Actual length	Min. estimate	Max. estimate	Item	Actual length	Min. estimate	Max. estimate
bug	5 mm	0.35 mm	15 cm	test tube	10 ml	1 ml	200 ml
chalk	8 cm	1 cm	15 cm	glass	100 ml	1 ml	1 l
book	46 cm	3 cm	90 cm	vase	300 ml	3 ml	2 l
table	70 cm	8 cm	1.30 m	carafe a)	500 ml	2 ml	2 l
room	3 m	2 m	6 m	carafe b)	1 l	500 ml	3.5 l
				pot	3.5 l	200 ml	10 l

Table 2: Maximum deviations from the real values

For almost all objects an underestimation of nearly 100 % can be found. Only one item of the lengths (room: 40 % deviation) and one item of the capacities (big carafe: 50 % deviation) show better values. Looking at the overestimations, a greater variety can be stated: Whereas the overestimations of the lengths differ by 80 to 100 % from the real values, the maximum deviations of the capacities range between 100 and nearly 2,000 %.

The same tendencies can be seen in the extremes of the experts' estimations, although the deviations are much smaller.

If we do not focus on the extremes, but on the means of deviations in the children's estimations, it can be stated, that on average the lengths were mostly underestimated, whereas the capacities were underestimated as well as overestimated.

Length overestimated: room (+2%).
Length underestimated: table (-6%), book (-17%), chalk (-20%), and bug (-30%).
Capacities overestimated: glass (+7%), small carafe (+20%), test tube (+68%).
Capacities underestimated: vase (-11%), big carafe (-15%), pot (-26%).

Looking at the means of the positive values of deviations, the estimations of lengths show a very uniform picture with the positive exception of the room: bug (M 37.8%; SD 27.3), chalk (M 31.8%; SD 22.0), book (M 32.2%; SD 23.1), table (M 31.0%;

SD 21.0), room (M 18.2%; SD 21.4). Perhaps the height of the room is a known value for a greater number of children. The values of the estimated capacities show much greater mean deviations as well as very high standard deviations: test tube (M 113.4%; SD 248.1), glass (M 93.6%; SD 95.0), vase (M 71.2%; SD 44.4), small carafe (M 42.2%; SD 38.2), big carafe (M 27.2%; SD 36.8), pot (M 41.1%; SD 25.1). Again, it may be that the carafes are better known than a test tube.

All in all, the results of the experts show less extreme deviations and were in total closer to the real values. But they are more likely to overestimate than to underestimate. Since only 17 experts participated, no means and standard deviations are given here.

THE DATA FROM THE PERSPECTIVE OF DIFFERENT CRITERIA

Overall application of different criteria

Tables 3 and 4 show the overall results of the children, and the experts. All estimations were accumulated and evaluated by the criteria mentioned above.

Criteria of ‘accuracy’	10 %	25 %	33 %	50 %	> 50%
Length (total # answers: 540)					
absolute	115	255	325	442	98
relative	21.3 %	47.2 %	60.2 %	81.9 %	18.1 %
Capacity (total # answers: 549)					
absolute	92	161	166	277	272
relative	16.8 %	29.3 %	30.2 %	50.5 %	49.5 %

Table 3: cumulated ‘accurate’ answers of the children using different criteria

The results of the children as well as of the experts show, that the estimation of lengths is easier than the estimation of capacities.

Criteria of ‘accuracy’	10 %	25 %	33 %	50 %	> 50%
length (total # answers: 85)					
absolute	48	76	78	81	4
relative	56.5 %	89.4 %	91.8 %	95.3 %	4.7 %
Capacity (total # answers: 81)					
absolute	24	45	50	64	17
relative	29.6 %	52.9 %	58.8 %	79.0 %	21.0 %

Table 4: Cumulated ‘accurate’ answers of the experts using different criteria

Concerning the different criteria, there is nearly no difference for the children's results between 25 % or 33 % in the capacity-condition, whereas it gives a good differentiation in the application to the estimated lengths and also to the experts' results, if this differentiation is necessary.

Ranges of deviations from the perspective of different groups of students

Table 5 shows the deviation-ranges of the estimation that were given by the best quarter and the best half of the children.

Deviations in estimating length			Deviations in estimating capacity		
Item	Best quarter	Best half	Item	Best quarter	Best half
bug	10 %	35 %	test tube	20 %	77 %
chalk	13 %	25 %	glass	48 %	85 %
book	11 %	30 %	vase	33 %	65 %
table	14 %	28 %	carafe a)	0 %	40 %
room	0 %	13 %	carafe b)	0 %	10 %
			pot	21 %	43 %

Table 5: Ranges of deviations of the best 25 % (50 %) estimations

Again, the results for the room and the carafes show that these estimations have been easy for at least the best quarter of students. It also becomes clear that the items differ in their difficulty especially in the capacity-condition.

DISCUSSION

Looking at the data and the application of the criteria, the following suggestions have to be discussed:

- It seems necessary to use different criteria for the evaluation of estimates in different measurement areas. The children and the experts gave better estimations for lengths than for capacities. The 17 experts also chose smaller ranges for lengths as adequate for evaluating children's estimations. But: Which ranges are adequate for which measurement area?
- A multi-step evaluation seems to be more adequate than a single-step one. But: How many steps should be differentiated? Is the number of steps different in different measurement areas?
- Since even our items differed in their difficulty, we seem to need different evaluations for them. But we are not sure yet, if there is a medium bandwidth in every measurement area in which it is easier to estimate. Do we have to define such bandwidths and use different criteria for evaluation if an item is in it or not?
- Last but not least: How many items have to be estimated to get a realistic

picture of a child's performance? How do we take the age of the child into account?

Nevertheless, the overall question remains, if the decision about an adequate estimate is a normative one or if it may be solved experimentally. But: Should a poor result get a good evaluation because it's the average of performance?

References

Bright, G. W. (1976). Estimation as part of learning to measure. In D. Nelson & R. E. Reys (Eds.), *Measurement in school mathematics* (pp. 87-104). Reston: NCTM.

Clayton, J. G. (1992). Estimation in schools. MPhil-thesis in Education. London: University of London. Retrieved from http://www.johngclayton.co.uk/website%20files/Output complete%20text.pdf [15-02-14].

Freudenthal, H. (1978). *Vorrede zu einer Wissenschaft vom Mathematikunterricht*. München: Oldenbourg.

Hildreth, D. J. (1980). *Estimation Strategy Use in Length and Area Measurement Tasks by Fifth and Seventh Grade Students.* Abstract retrieved from PhDTree http://phdtree.org/24304583-estimation-strategy-use-in-length-and-area-measurement-tasks-by-fifth-and-seventh-grade-students/ [22-09-14].

Hogan, T. P. & Brezinski, K. L. (2003). Quantitative estimation: One, two, or three abilities? *Mathematical Thinking and Learning, 5*(4), 259-280.

Huang, H.-M. E. (2014). Investigating children's ability to solve measurement estimation problems. In S. Oesterle; P. Liljedahl,; C. Nicol & D. Allan (Eds.), *Proceedings of the Joint Meeting of PME 38 and PME-NA 36*, Vol. 3, pp. 353-360. Vancouver, Canada: PME.

Jones, M. G.; Forrester, J. H.; Gardner, G. E.; Andre, T. & Taylor, A. R. (2012). Students' accuracy of measurement estimation: context, units, and logical thinking. *School Science and Mathematics, 112*(3), 171-178.

Lörcher, G. A. (2000). Zur Entwicklung der Zahlvorstellung. In M. Neubrand (Ed.), *Beiträge zum Mathematikunterricht: Vorträge auf der 34. Tagung für Didaktik der Mathematik 28.2.-3.3.2000*, pp. 402-405.

O'Daffer, P. (1979). A case and techniques for estimation: Estimation experiences in elementary school mathematics – essential not extra! *Arithmetic Teacher, 26*(6), 46-51.

Ruwisch, S. & Heid, M. (2015). Children's use of strategies in estimating length and volume. Poster presentation. Hobart, Australia: PME. [for submisson 15-03-01].

Siegel, A. W.; Goldsmith, L. T. & Madson, C. R. (1982). skill in estimation problems of extent and numerosity. *Journal for Research in Mathematics Education, 13*(3), 211-232.

Sowder, J. (1992). Estimation and Number Sense. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 371-389). Reston: NCTM.

Swan, M. & Jones, O. E. (1980). Comparison of students' perceptions of distance, weight, height, area, and temperature. *Science Education, 64*(3), 297-307.