WHICH ESTIMATION SITUATIONS ARE RELEVANT FOR A VALID ASSESSMENT OF MEASUREMENT ESTIMATION SKILLS

Heinze, Aiso; Weiher, Dana Farina; Huang, Hsin-Mei; Ruwisch, Silke

Published in:
Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education.

Publication date:
2018

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

PME 42

July 3-8, 2018 --- Umeå, Sweden

Proceedings

Of the 42nd Conference of the International Group for the Psychology of Mathematics Education

Editors: Ewa Bergqvist, Magnus Österholm, Carina Granberg, and Lovisa Sumpter

Volume 3

Research Reports Ham – Pi

Proceedings of the 42nd Conference
of the International Group
for the Psychology of Mathematics Education

Umeå, Sweden
July 3-8, 2018

Editors

Ewa Bergqvist
Magnus Österholm
Carina Granberg
Lovisa Sumpter

Volume 3

Research Reports Ham – Pi

The 42nd Annual Meeting of the International Group for the Psychology of Mathematics Education

PME 42

July 3-8, 2018 --- Umeå, Sweden

www.pme42.se

Cite as:

Bergqvist, E., Österholm, M., Granberg, C., & Sumpter, L. (Eds.). (2018). *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3). Umeå, Sweden: PME.

Website: http://www.pme42.se

The proceedings are also available via http://www.igpme.org

ISSN: 0771-100X

ISBN (volume 3): 978-91-7601-904-7

Logo Design: Catarina Rudälv and Amanda Rudälv

Printed by CityPrint i Norr AB, Umeå

WHICH ESTIMATION SITUATIONS ARE RELEVANT FOR A VALID ASSESSMENT OF MEASUREMENT ESTIMATION SKILLS?

Aiso Heinze[1], Dana Farina Weiher[2], Hsin-Mei Huang[3], and Silke Ruwisch[2]

[1]IPN – Leibniz Institute for Science and Mathematics Education Kiel, Germany
[2]Leuphana University Lüneburg, Germany; [3]University of Taipei, Taiwan

Measurement estimation skills are of significant importance for everyday life. In the last decades a lot of research results were generated describing students' estimation skills and strategies. Surprisingly, little attention has been paid to the basic question which types of situations are relevant for a valid conceptualization and operationalization of measurement estimation skills. Some studies refer to the basic structure of estimation conditions described by Bright (1976) whereas others ignore this question, though it is central to ensure validity of the empirical data. Following validity criteria and based on existing empirical findings on estimation strategies, we developed a comprehensive model of measurement estimation situations. This model provides a basis for the development of valid tests on measurement estimation skills as well as for the development of learning environments.

THEORETICAL BACKGROUND

Skills to estimate the attributes of objects (e.g., length, area) are of significant importance for everyday life as well as for professional expertise in various professions (Jones, Taylor, & Broadwell, 2009) and can be considered as a basis of measurement skills (Joram, Subrahmanyam, & Gelman, 1998). To date, research provides a lot of information on individuals' measurement estimation process, strategies and performance (e.g., Siegel, Goldsmith, & Madson, 1982; Joram et al., 2005). Moreover, empirical findings show that the teaching of estimation strategies is possible (e.g., Hildreth, 1983) and improves the accuracy in students' measurement estimation (e.g., Joram et al., 2005; Jones et al., 2009).

Most of the empirical studies used items representing specific estimation situations to collect data on estimation skills or strategies. Surprisingly, many studies did not address the question which types of estimation situations are relevant to elicit the skills or strategies aimed for from the considered students or adults. Ignoring the choice of estimation situations might result in a serious threat of validity of empirical data and its interpretations. We elaborate on this problem and suggest a comprehensive model of measurement estimation situations which satisfies validity criteria and integrates current research results on measurement estimation skills. For reasons of simplicity, we restrict our presentation on length estimation, though our model can probably be adopted for other attributes like area and volume as well.

2018. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.). *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 67-74). Umeå, Sweden: PME.

Measurement estimation and measurement estimation strategies

In the sense of Bright (1976) and other researchers, we consider measurement estimation as a mental process of determining a measurement for an attribute of an object without the aid of measurement tools. Central to this process is that the use of measure units happens mentally while other aids like benchmarks or body parts might be used as additional tools (e.g., estimate the length of a pencil in cm as a benchmark and then determine the width of the table by repeated use of the pencil as a tool).

Empirical research yields that children and adults mostly show a poor measurement estimation performance, that individual length estimation is in general more accurate than the estimation of area, volume, or weight and that estimation performance of students increase with grade (e.g., Siegel et al., 1982; Sowder, 1992; Joram et al., 1998). In order to understand estimation performance, Siegel et al. (1982) analyzed students' estimation processes and developed a process model describing the individual estimation process. This model particularly emphasizes the role of different estimation strategies in the estimation process and subsequent research provided evidence that the use of strategies predicts estimation performance (Joram et al., 2005; Jones et al., 2009; Huang, 2015). The most important estimation strategies (e.g., Siegel et al., 1982; Hildreth, 1983; Joram et al., 1998) are

- *unit iteration* as a mentally conducted measurement by a segmentation of the to-be-estimated object (TBEO) based on a given standard or non-standard unit and subsequent counting of the segments;
- *benchmark comparison* (or reference-point strategy) as a mental comparison of the TBEO with a distance represented by a benchmark or the sequence of the same benchmarks where the length of the benchmark is known or can be estimated;
- *decomposition/recomposition* as a process of mental decomposition of the TBEO into smaller parts, estimation of the length of each part by using one of the previously mentioned strategies and adding the estimates of all parts.

In addition to the estimation strategies, research points to further components of estimation skills. As the descriptions of the strategies make clear, domain-specific knowledge obviously plays an important role. This encompasses, for example, measurement knowledge related to the standards units (i.e., mm, cm, m, km and their interrelations in case of the metric system) as well as knowledge on benchmarks in a twofold way (e.g., knowledge on the approximate width of an A4 sheet which is given as a possible benchmark; knowledge of a suitable object that can be used as a benchmark for 10 cm). Besides these knowledge components, specific cognitive abilities contribute significantly to estimation performance. Models of estimation processes from cognitive psychology assume that (1) the TBEO is represented in the working memory, (2) this representation is estimated based on estimation strategies and information retrieved from the long-term memory and (3) the estimated length is finally checked in a monitoring process (e.g., D'Aniello, Castelnuovo, & Scarpina, 2015). From this follows that beyond knowledge on measurement, benchmarks and estimation strategies (stored as information in the long-term memory) individual working memory capacities play an important role.

As we already mentioned before, the relevant knowledge for the estimation process can be acquired in effective learning environments. This means in particular that students educated in different learning environments based on different curricula might possess different knowledge (e.g., different estimation strategies). Such differences become obvious when we consider students from two countries with different cultures, educational traditions and curricula. Differences in the learning context probably result in different benchmark knowledge since benchmarks are influenced by the cultural context. Differences might also occur in how students learn to implement estimation strategies. For example, some curricula emphasize the use of body parts as benchmarks for touchable TBEOs as suggested in Jones et al. (2009), other curricula may emphasize the strictly mental use of benchmarks to estimate imagined TBEOs.

Measurement estimation situations as a basis for research on estimation skills

In general, measurement estimation skills are inferred from the performance of an individual generating an accurate estimate for a required measurement of a given attribute of an object. To assess these estimation skills, individuals are asked to solve various estimation items. Such items represent estimation situations which can differ substantially and therefore might influence the estimation performance. For example, Pike and Forrester (1997) administered items representing estimation situations in a story context (ladybirds amidst a rainfall) and in a stereotypical mathematics textbook context. –It turned out that students' estimation performance was better in the mathematics textbook context – a phenomenon which is probably caused by the specific type of mathematical tasks and activities implemented in mathematics textbooks and the mathematics classroom. However, even when restricting items to one specific context (e.g., real life context), in each estimation situation there are characteristics which must be understood by students before performing an estimation process and which thus might influence the item difficulty. Accordingly, a thorough analysis of estimation situations implemented in test or interview items is necessary to ensure validity of the empirical data and inferred results. Surprisingly, in many published studies this challenge is addressed neither explicitly nor implicitly. In the following we shortly present three examples retrieved from the literature: first, the model of Bright (1976) who explicitly addressed this question and to which other articles refer (e.g., Sowder, 1992) and then the descriptions from the studies of Jones et al. (2009) and Siegel et al. (1982).

Model of Bright (1976)

In Bright (1976), eight types of estimation situations are described each as combination of three independent characteristics: (1) the object or the measurement is specified, (2) the TBEO is physically present/given or not, and (3) the unit of measurement is given or not. The eight situation types are divided into two parts, first the situation types where the object is specified (A) and second the types where the measurement is specified (B). The four situation types of class A are the usual estimation situations where students have to estimate a measurement for an attribute of an object, whereas the four situation types of class B are mainly interesting for instructional purposes (to support students in generating benchmark knowledge). It is mentioned (Bright, 1976, p. 90) that further subdivisions of the situation types could be made.

Structure in Siegel, Goldsmith, and Madson (1982)

In their article, Siegel and colleagues present an estimation process model based on findings of an interview study. The study relies on specific items suggesting the use of specific estimation strategies in order to elicit the cognitive processes of interest. Hence, there is an implicit model of different types of estimation situations structured by estimation strategies. Siegel et al. (1982) distinguish two problem types related to length measurement: benchmark problems and decomposition problems. However, in this case benchmark problems do not explicitly ask for the use of a benchmark or explicitly mention a benchmark. Instead, it is assumed that problems like "How long is a piece of manuscript paper?" are solved by the benchmark strategy. In contrast, the presented decomposition items explicitly describe decompositions ("If you took these cooking utensils and laid them end to end, how far would they reach?"). For both problem types the TBEOs were presented physically or by photographs. From the article it does not become clear whether the children and adults were allowed to touch the TBEOs during the estimation process.

Model of Jones, Taylor, and Broadwell (2009)

The article of Jones and colleagues on the use of body parts in the estimation process describes the Linear Measurement Assessment (LMA) which they used to test linear measurement estimation skills. The LMA is based on a model with five dimensions representing different types of length estimation situations (Jones et al., 2009, p. 1502): (1) estimating the length of an object while viewing the object; (2) naming an object from memory for different metric sizes; (3) estimating the lengths of large objects like a building; (4) metric estimation of objects that students can touch or distances they can pace; (5) using body parts as an aid to measure different objects. Analyzing the types of estimation situations, it turns out that different aspects like size or presence of the TBEO as well as the option to touch the TBEO are varied. Moreover, the situation type (2) is similar to one of those in Bright's (1976) situation types of class B.

Summarizing the state of the art, it turns out that many studies in the field of measurement estimation skills do not explicitly elaborate on the choice of estimation situations for their data collections. Some studies refer to the model of Bright (1976), others provide own criteria for structuring the estimation situations. As Bright (1976) mentioned, his model can be refined and analyzing the other existing models, it turns out that specific types of estimation situations are not distinguished (e.g., situations in which a touchable benchmark is given or a representation of the TBEO's length can be constructed by drawing a line). However, it is not clear what grain-size is relevant for research and educational practice in the field of length estimation skills.

RESEARCH GOAL AND RESEARCH APPROACH

Based on the previously presented theoretical background, we elaborate on the question what types of estimation situations are relevant for the assessment of estimation skills. As mentioned in the beginning, we restrict our presentation to length estimation. Our goal is to develop a comprehensive model on types of length estimation situations which ensures validity in case of assessments of length estimation skills. To establish validity, we follow the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014). In chapter 1, the standards provide five sources of evidence which can contribute to validity: evidence based on (1) content, (2) response processes, (3) internal structure, (4) relations to other variables, and (5) consequences of the interpretation of results.

As presented in the previous section, empirical research yields that the following aspects are relevant factors for length estimation skills: knowledge in estimation strategies, measurement knowledge, benchmark knowledge, working memory capacity as well as context factors like culture (influences which estimation strategies are emphasized in the mathematics classroom) and characteristics of estimation situation in real life (in contrast to estimation situations in mathematics textbooks). These factors for length estimation skills contribute to the five sources of evidence for validity as follows (cf. AERA, APA, & NCME, 2014):

1. *Evidence based on content* asks for the relation between the construct (length estimation skill) and the requirements in the estimation situations. Hence, the situations must cover all relevant aspects of estimation situations in real life.
2. *Evidence based on response processes* asks for a fit between the construct (length estimation skill) and the observed performance in estimation situations. It must be ensured that the performance of mastering the estimation situations is based on the components (knowledge, working memory capacity) of estimation skills.
3. *Evidence based on the internal structure* asks for a fit between the structure of the construct (length estimation skill) and the performance in different types of estimation situations. Here, the influence of culture, educational traditions and curricula might play a role because the teaching of different estimation strategies can yield different performance profiles provided suitable estimation situations are represented by the items.
4. *Evidence based on relations to other variables* asks for correlations between the estimation performance and other variables. For example, performance should improve with the increase of benchmark knowledge or the grade of the students.
5. *Evidence based on consequences of the interpretation of results* means in particular to exclude unintended interpretations of the empirical results caused by construct irrelevant components or by construct underrepresentation. For example, culture might influence benchmark knowledge so that estimation situa-

tions with given benchmarks can be problematic in cross-cultural studies. Construct underrepresentation might occur if estimation situations are restricted to certain types (e.g. only touchable TBEOs) so that specific components of estimation skills are not required.

RESULTS

Based on the framework described by the criteria 1-5 in the previous section and existing models of estimation situations presented in the theoretical background, we identified six characteristics of estimation situations. The six characteristics are mostly pairwise independent; however, some combinations do not make sense in real life or even cannot occur so that these situations are excluded. Figure 1 gives an overview of the 72 estimation situations as compact tree diagram. It should be mentioned that there is a seventh characteristic which is not displayed in the tree diagram and which is related to the magnitude of the TBEO. Since length estimation of large objects is more challenging than length estimation of small objects (e.g., Jones et al., 2009), each estimation situation should be considered for small, medium, large, and huge objects.

There are manifold relations of the presented characteristics of estimation situations to the five validity criteria 1-5 developed in the previous section. Due to space limitations, we cannot explain and argue in detail how each characteristic of our model is related to the five criteria so that we restrict to some exemplary aspects.

The seven characteristics represent relevant aspects of estimation situations in real life and thus contribute to the first criterion (content validity). In real life, the TBEOs can be (i) physically present or not and (ii) touchable or not, (iii) there might be the opportunity to construct a representation of the same length (e.g., by drawing a line) and (iv) it can happen that no benchmark is given or that a specific benchmark (or more than one) is mentioned, is visible in real size, or is even visible in real size and touchable. Moreover, there are situations which (v) ask for estimates in standard units (e.g. metric units like cm) or in non-standard units (e.g., room width in number of floor tiles). If a benchmark is given, it can happen that (vi) its length is provided or not. Finally, (vii) in real life the TBEOs can vary in their length from small to huge.

In addition to the first criterion, the characteristics satisfy the other four criteria of validity. For example, it makes a strong difference for the cognitive load of the working memory if a TBEO is touchable or not or if a benchmark is visible in real size or not. Estimation processes are more challenging if individuals must imagine the TBEO and the benchmark because external representations are not available. Due to the high cognitive load, it can happen that only efficient estimation strategies can be processed and persons who only know complex strategies will show a lower performance in these situations. Hence, taking into account these characteristics of estimation situations contributes to validity corresponding to criteria 2-5 because varying performance in different types of estimation situations can be distinguished. As a second example, we want to mention that the seven characteristics allow the distinction of estimation situations which do or do not support specific estimation strategies. If a TBEO is touchable, the benchmark comparison strategy is much easier because a person can directly or indirectly use body parts as benchmarks (Jones et al., 2009). If representations can be constructed (e.g., by drawing a line of the length of the TBEO), this might support the decomposition strategy in case of medium sized TBEOs. The application of estimation strategies is associated with different knowledge components that depend on previous learning experiences. Hence, the seven characteristics allow distinguishing estimation situations which ensure on the one hand that persons with different knowledge on estimation strategies show different estimation performances (relevant for criteria 2-4) and avoid on the other hand that the construct length estimation skill is not underrepresented in the assessment (relevant for criterion 5).

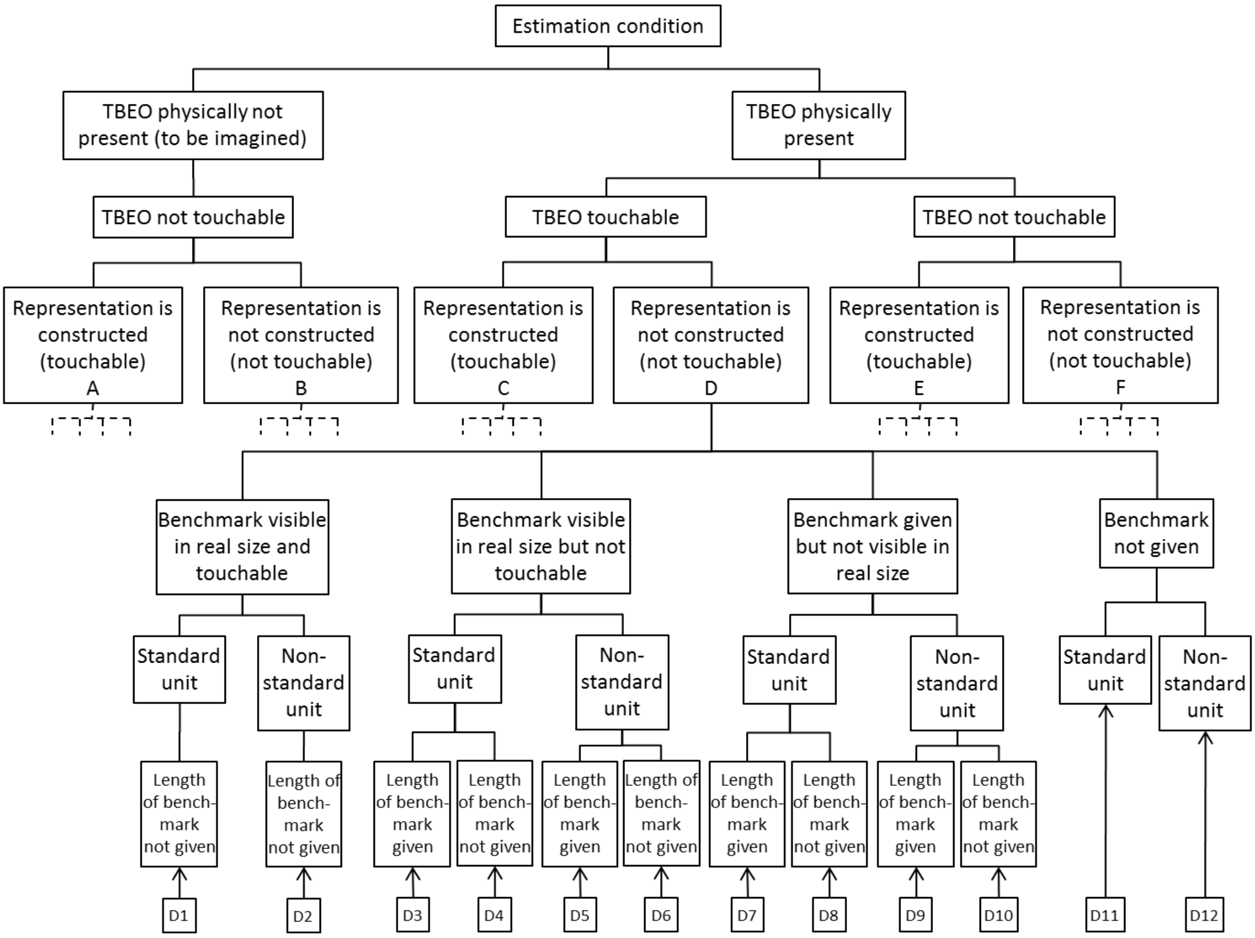

Figure 1: Model of 72 types of length estimation situations (12 types for cases A-F).

DISCUSSION

In this contribution we argued that in empirical research on measurement estimation skills, the choice of estimation situations is of great importance when collecting data. A low variation in the characteristics of estimation situations might result in a serious threat of validity of empirical data and the interpretations of the results. Surprisingly, many articles of empirical studies do not explicitly report how they have chosen the estimation situations for their assessment instruments and a reconstruction of this information is not possible. Only some studies refer to the existing basic model of Bright (1976) or give an explicit or implicit description of an own model.

For our analysis we adopted aspects of validity from AERA, APA, and NCME (2014) and combined these with the current state of research on relevant aspects of length estimation skills. The resulting five criteria for validity allowed developing a comprehensive model of types of estimation situations (Figure 1). In comparison to the models of Bright (1976), Siegel et al. (1982) and Jones et al. (2009), our model is much more detailed. It allows distinguishing more types of estimation situations and therefore gives opportunities for a more detailed analysis of estimation skills.

Currently, our arguments are purely theoretical. Hence, it is an empirical question how fine-grained the estimation situations must be mirrored by items in empirical studies. Depending on the respective research goal and the kind of data that should be collected, a coarser model with fewer estimation situations might be sufficient to assess length estimation skills. In such cases the model of types of estimation situations in Figure 1 can serve as an ideal to check for validity.

References

AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

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

D'Aniello, G. E., Castelnuovo, G., & Scarpina, F. (2015). Could cognitive estimation ability be a measure of cognitive reserve? *Frontiers in Psychology, 6,* 1-4.

Hildreth, D. J. (1983). The use of strategies in estimating measurements. *Arithmetic Teachers, 30*(5), 50-54.

Huang, H.-M. E. (2015). Children's Performance in Estimating the Measurement of Daily Objects. In K. Beswick, T. Muir & J. Wells (Eds.). *Proceedings of 39th Psychology of Mathematics Education conference,* (Vol. 3, pp. 73-80). Hobart, Australia: PME.

Jones, G., Taylor, A., & Broadwell, B. (2009). Estimating linear size and scale: Body rulers. *International Journal of Science Education, 31*(11), 1495-1509.

Joram, E., Gabriele, A. J., Bertheau, M., Gelman, R., & Subrahmanyam, K. (2005). Children's use of the reference point strategy for measurement estimation. *Journal for Research in Mathematics Education, 36(1),* 4-23.

Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Journal of Educational Review*, 6, 413-449.

Pike, C. D., & Forrester, M. A. (1997). The influence of number-sense on children's ability to estimate measures. *Educational Psychology, 17*(4), 483-500.

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.