Modelling the Complexity of Measurement Estimation Situations – A Theoretical Framework for the Estimation of Lengths

DANA FARINA WEIHER, LÜNEBURG; SILKE RUWISCH, LÜNEBURG; HSIN-MEI E. HUANG, TAIPEI; JESSICA HOTH, KIEL & AISO HEINZE, KIEL

Zusammenfassung*: Schätzfähigkeiten sind alltagsrelevant und werden mit verschiedenen Forschungsinteressen untersucht. Trotz der Verwendung unterschiedlicher Schätzsituationen setzen sich die Studien kaum mit Fragen der Aufgabeneigenschaften auseinander. Dieser Beitrag stellt ein theoretisches Modell der Aufgabenkonstruktion vor. Die wichtigsten Merkmale von Aufgaben zur Längenschätzung werden durch Analyse der Literatur herausgearbeitet. Im Anschluss werden diese Elemente systematisch variiert, kombiniert, re-analysiert und reduziert. Das Ergebnis ist ein Modell mit 228 sinnvollen Situationen für die Längenschätzung, das auch als analytisches Instrument für den Vergleich von Studien dienen kann.*

Abstract: *Abilities in estimation are important for everyday life, and investigated with different research interests. Although quite different estimation situations were used, studies usually did not address the question of task characteristics systematically. This paper presents a theoretical model that may serve as a basis for the task design in length estimation. First, key features of length estimation tasks were extracted by analyzing the literature. Second, these elements are varied systematically. Third, they are combined, reanalyzed, and reduced. This procedure results in a model of 228 meaningful situations for length estimation. The model may also serve as an analytical tool for comparison of studies.*

1 Estimation as a Complex Cognitive Process and the Question of Validity

Estimation is seen as a highly important ability for everyday life, in mathematics education literature (Forrester et al., 1990; O'Daffer, 1979; Sowder, 1992) as well as in cognitive psychology (Della Sala et al., 2003a; Harel et al., 2007; Liss et al., 2000; MacPherson et al., 2014). Like problem solving, estimation is considered a complex cognitive process. Therefore, abilities in estimation, especially the accuracy of the estimation results, serve for many different research interests using various estimation tasks and research designs.

Cognitive psychologists mainly use paper-pencil estimation tests as an indicator of the executive functions in diagnosing different cognitive deficits. They also serve for the investigation of the relation between estimation abilities and other cognitive functions like semantic knowledge or working memory (e.g., Brand et al., 2003; Bullard et al., 2004; D'Aniello et al., 2015; Della Sala et al., 2003b; MacPherson et al., 2014, Shallice, & Evans, 1978).

Whereas cognitive psychologists treat estimation as one construct, three kinds of estimation are differentiated in mathematics education: numerosity estimation, computational estimation and measurement estimation (O'Daffer, 1979; Sowder, 1992). With respect to Hogan and Brezinski (2003), abilities in computational estimation are separable from the two others. By re-analyzing the estimation tests in cognitive psychology with respect to this differentiation, it can be seen that computational estimation tasks are not included, but all tests used different mixtures of tasks for numerosity estimation as well as for measurement estimation.

When focusing on measurement estimation itself, it can be stated that most psychological tests combined different measurement areas in one paper-pencil-test. All studies contained the estimation of lengths, associated with tasks from one or the other measurement area: the estimation of weight, time, money or speed (e.g., Axelrod, & Millis, 1994; Bullard et al., 2004; Della Sala et al., 2003b; Goldstein et al., 1996; MacPherson et al., 2014; Mendez et al., 1998; Shallice, & Evans, 1978). Only few tests contained the estimation of capacity, area (Della Sala et al., 2003b) or temperature (Axelrod, & Millis, 1994) and – as far as we know – none asked for the estimation of volume. This situation leads to the fact that most estimation tests used in cognitive psychology do not include an equal number of each kind of estimation or measure. This might be reasonable, because the researchers were interested in the executive functions involved in any complex cognitive process. Nevertheless, if being interested in estimation abilities themselves, these tests seem to be too unspecific. From an analytical point of view, they also seem problematic, because the dimensionality of measurement estimation abilities themselves, especially the abilities to estimate different attributes, is not clear yet (Heinze et al., 2018; Hogan, & Brezinski, 2003).

Concerning measurement estimation in mathematics education, the studies were also conducted with different aims and included different tasks. As in cognitive psychological tests, most studies in mathematics education combined different measurement areas in one inquiry (Corle, 1963; Forrester et al., 1990; Heid, 2018; Hildreth, 1980; Hogan, & Brezinski, 2003; Huang, 2014; Pike, & Forrester, 1997; Siegel et al., 1982; Swan, & Jones, 1980). Only few studies focused on only one measurement area; if so, they concentrated on length measurement (Desli, & Giakoumi, 2017; Jones et al, 2009, 2012; Joram et al., 2005).

Apart from the question of the involvement of different measurement areas, it is also surprising that many studies did not address the question which characteristics of the estimation tasks are relevant to elicit the abilities or strategies aimed for. Although researchers assumed task characteristics being responsible for example for strategy choices (e.g., Siegel et al., 1982), there is still a lack of a systematic analysis of variables in situations in which participants are asked to estimate measures (Joram et al., 1998; Sowder, 1992).

Not paying enough attention to the question of situational conditions in measurement estimation affordances might be one reason for the unclear validity of the data in the field of measurement estimation research.

2 Aim of the Paper and Line of Argument

The aim of this article is to develop and present a theoretical model for possible measurement estimation tasks.[1] The underlying intention is to provide a foundation for better comparisons between studies in the field of measurement estimation and thereby increase the transparency. Our considerations will be concentrated on lengths, though the model might probably be adapted to other attributes as well.

According to Bright (1976) and in mathematics education widely accepted (e.g., Hildreth, 1980; Joram, 2003; Sowder, 1992), two task types in measurement estimation can be distinguished:

- The size of an object is to be estimated: within this type of task the to-be-estimated-object (further named as TBEO) and the unit can be physical pre- or absent;
- A proper object for a given measure is to be named: within this type of task, possible objects and the unit can be given or not.

Finally, eight different types of measurement estimation tasks were constructed in this way. This typology of Bright (1976) served as a starting point for our development of a framework for possible length estimation situations presented in this paper.[2]

Before presenting our suggestions for such a fundamental framework of length estimation tasks in section 4 and 5, we will take a closer look to the mathematics education literature dealing with measurement estimation. The main interest of our analysis was to get information about task characteristics which should be involved in our model. By doing so, we came to the realization that some terms and statements are used with different underlying meanings, so the results of those studies may also not be as alike as they seem to. Therefore, section 3 focuses on two goals: We present the information we got from these studies and discuss how the results suffer from not being comparable so easily because of differences in their underlying structure. In particular, they differ in their definitions of terms and the characteristics of the tasks they used.

With the help of the resulting model, the concept of length estimation can be operationalized more explicitly, which hopefully will lead to more valid and comparable studies about measurement estimation.

3 Length Estimation in Mathematics Education Literature

Focusing on measurement estimation it seems plausible to start with lengths. Lengths are the most fundamental measures, in daily life as well as in mathematics learning. They also serve as the fundament for other geometrical measures, namely area and volume measurement, and therefore gained the most interest in studies.

In analyzing the mathematics education literature with respect to length estimation, we start with remarks to the definition. In looking to two prominent examples we want to draw attention to the differences and effects of – sometimes even implicit – underlying definitions.

In the following section, estimation strategies are presented with the same focus: Whereas 'benchmarks' and 'unit iteration' are seen as crucial elements of successful estimation processes by all authors, we will also show by examples the differences in using these terms for different behavior and therefore not being comparable easily.

The final section presents information about task characteristics in detail: we sum up what is known about the impact of special characteristics of the tasks used in length estimation. Again, we distract important elements and discuss differences between studies with the 'same' task design.

Whereas the information from the last paragraph directly influenced our suggestions for a theoretical model, the first two paragraphs had more implicit impact to it.

3.1 The Question of Definition

Mitchell et al. (1999) defined estimation "[a]s a process whereby one approximates, through rough calculations, the worth, size, or amount of an object or quantity that is present in a given situation. The approximation, or estimate, is a value that is deemed close enough to the exact value or measurement to answer the question being posed" (p. 9). Bright (1976) defined measurement estimation as a "process of arriving at a measurement or measure without the aid of measuring tools. It's a mental process" (p. 89). Both definitions are widely accepted and referred to by the mathematics education community, even though they differ in crucial aspects. Although the former definition is broader in the sense of included kinds of estimations – numerosity, computational and measurement estimation – it is restricted in the use of strategies and situational contexts: rough calculation of objects or quantities that are present. However, the object, which length should be estimated, can also be absent, and besides rough calculation, several other strategies can be used (Desli, & Giakoumi, 2017; Heid, 2018; Hildreth, 1980, 1983; Joram et al., 1998; Siegel at al., 1982). The latter definition also is restricted in two ways: First, it is focused on measurement estimation as the only kind, and second, the approximating process is restricted to mental processes without any helping tool.

Whereas some authors see this restriction to mental processes very strictly concerning all aspects of the estimation process, others use it in a broader way: Central to the estimation process is always the strictly mental use of measure units – so no rulers or other measurement tools are allowed – whereas other aids like body parts or other benchmarks might be used as supporting tools. If, for example, being asked to estimate the width of the table, the former would not allow any 'action', whereas in the latter sense the children may estimate the length of their pencil as a benchmark and then determine the width of the table by iterated use of the pencil's length.

In this article and our model, we restrict ourselves to length measurement, suggesting that it might be transferable to other measurement areas as well. Nevertheless, this is not done yet and may also be an empirical question. Our framework suggests for building on the broader definition: Following this idea, we define length estimation as a measurement process which necessarily includes at least one mental step of assessing (a) the size of the TBEO or (b) the size of a part of the TBEO without measuring with a standardized measurement instrument.

3.2 The Question of the Impact of Strategies

Besides the question of defining length estimation itself, the elaboration of different solution strategies is the most discussed issue in mathematics education literature dealing with measurement estimation. In the literature (Desli, & Giakoumi, 2017; Hildreth, 1980, 1983; Heid, 2018; Joram et al., 1998, 2005), a great variety of names for estimation strategies can be found. We will not discuss all of them here in detail (for an overview, see e.g., Joram et al. (1998), Heid (2018); Weiher, & Ruwisch (2018)), but clarify our understanding and extract those elements which may have an impact on our model.

All measurement estimation strategies are based on comparisons between the TBEO and an object whose measure is known and can be used as a referential object, as a mental measurement unit.[3] These referential objects are called benchmarks and can be seen as a mental 'tool' for measurement estimation (Joram et al., 2005).

Any object can be used as a benchmark. In the literature (Desli, & Giakoumi, 2017; Hildreth, 1980, 1983; Joram et al., 1998, 2005) it is often differentiated whether the object is a standard unit or standard measurement instrument or if it is an everyday object familiar to the estimator. In most cases, only the latter is called 'benchmark strategy' whereas the former is named 'unit iteration'.

Using these terms in such a restricted manner seems problematic for the following reasons: First, standard units cannot be visualized without objects representing them (e.g., one centimeter being mentally represented by the distance between two longer lines – the cm-marks – on the ruler). Consequently, there is no logical difference between a mental representation of an everyday object or a standard unit. Second, both kinds of objects can be used for a mental measurement process, which should be named by 'unit iteration', although in the case of an everyday object the unit may be a nonstandard one. The process of using any unit and iterating it mentally is not restricted to standard units. Third, a mental comparison without any iteration can also be done with both kinds of objects, so the benchmark strategy is not restricted to everyday objects but can also be done with representations of a standard unit.

In re-analyzing the literature (Desli, & Giakoumi, 2017; Hildreth, 1980, 1983; Joram et al., 1998, 2005), another problem occurred. It seems that children who reported having used the strategy 'unit iteration' only used a vague mental image of what they think one cm, one inch etc. looks like without even iterating it mentally. In contrast, children who reported having used any daily object were able to describe their mental measurement process – most iterated it – in more detail. Because of the confusing signification of the terms 'benchmark strategy' and 'unit iteration', we suggest using the terms 'direct mental comparison' and 'indirect mental comparison' instead (Heid, 2018; Weiher, & Ruwisch, 2018):

- Direct mental comparison: If the TBEO and the referential object have nearly similar sizes, a direct comparison of the TBEO with the visually imagined size of the referential object may lead to an answer.
- Indirect mental comparison: If the TBEO and the referential object differ in their sizes, the comparison is more difficult.
  If the referential object is smaller than the TBEO, a mental unit iteration process can be worked out that uses the referential object as a unit to measure up the TBEO. If the referential object is bigger than the TBEO, it has to be mentally subdivided – for example into halves or quarters – until it fits nearly the length of the TBEO.

Apart from these main strategies of direct and indirect mental comparison, some other strategies are mentioned in the literature: One group of strategies is based on the idea of restructuring the estimation task. As such a preceding strategy, decomposition/recomposition (e.g., Siegel et al., 1982) means to decompose the TBEO into smaller parts for which benchmarks are available. One of those parts will be estimated, and this estimation result has to be multiplied by the number of parts (recomposed). If the TBEO is not subdivided into regular parts, every resulting part has to be estimated and the resulting estimates have to be added. The strategy decomposition/recomposition on its own is not leading to an estimation result: to get the estimate for one (each) part a comparison with benchmarks has to be done before the recomposition of the results (Siegel et al., 1982; Weiher, & Ruwisch, 2018).

Concerning the impact of different strategies on the estimation itself, especially on the accuracy of the estimation result, the statement of Joram et al. (2005) still meets the situation:

> Although mathematics educators have assumed a beneficial role for the use of strategies in estimating measurements, very few studies have demonstrated that use of these strategies is related to greater estimation accuracy (p.7).

In addition to the question of correlations between strategy use and estimation accuracy, Joram et al. (2005) showed that instruction in the use of strategies can be successful. They instructed two classes of 3rd graders for six lessons in using either the benchmark strategy – mentally imagine an everyday object or mentally iterating it – or a guess and check procedure solving length estimation tasks.

Joram et al. (2005) reported that only four students (9 %) used the benchmark strategy spontaneously during the pretest, whereas 27 students (63 %) mentally iterated a standard – customary – unit and twelve students (28 %) did not show any strategy. In the posttest 63 % of the students in the strategy-group used the benchmark strategy, whereas all others except one boy used the unit iteration strategy. 14 % of the students in the guess-and-check-group also showed the use of the benchmark strategy, while 55 % used unit iteration and still 32 % did not show any strategy. Joram et al. (2005) also reported that the accuracy in the estimates of students using the benchmark strategy compared to those not using it was significantly higher. Especially the smaller standard deviations showed that nearly all students estimated quite accurately, whereas in the other groups, students showed a wide spread of deviations from the real value in the estimates. The authors conclude that nearly all children who use the benchmark strategy will improve their estimation accuracy.

Since the authors neither discussed 'unit iteration' in detail nor reported differences between pre- and posttest of those students using this strategy solely, there is no information if these children also improved their accuracy in length estimation. Nevertheless, it seems our remarks on 'unit iteration' in the former paragraph also refer to this study.

Most studies did not investigate correlations between the use of strategies in length estimation and other characteristics. Due to this incompleteness, the following results seem almost arbitrary, but represent all relevant statements from the literature that are known to us. Hildreth (1980, 1983) showed that through instruction, the use of appropriate strategies of 5th and 7th graders increased and the use of inappropriate strategies decreased. In a study of Desli and Giakoumi (2017), 3rd and 5th graders showed a much broader variety of strategies when estimating length in standard units than in nonstandard units. In the latter case, mental unit iteration was almost exclusively used, and also the most observed strategy in dealing with standard units. Whereas none of the strategies correlated significantly with success when estimating in standard units, Desli and Giakoumi (2017) found positive correlations between success and the unit iteration strategy as well as the benchmark strategy and negative correlations between success and idiosyncratic responses in estimation situations with nonstandard units. According to Heid (2018), direct mental comparisons led to more accurate estimates than indirect mental comparisons (although the estimation accuracy was rather low overall).

Especially the differences between estimating in standard and nonstandard units, shown by Desli and Giakoumi (2017), had the impact on our model to consider this variation systematically.

3.3 The Question of Task Characteristics

If task characteristics are investigated in mathematics education studies, three different foci can be differentiated: The characteristics of the TBEO can be varied, the unit, in which the estimate of the TBEO should be given, can be varied, and other conditions, like possible benchmarks, allowed actions, or other 'contextual variables' can be varied as well.

Concerning the *characteristics of the TBEO*, Desli and Giokoumi (2017) recently conducted a study with primary students and their length estimation abilities (3rd and 5th graders). They focused on the influence of different task characteristics, which were already investigated by Jones et al. (2012) with older children. Concerning the TBEO they varied the following characteristics: small (up to 30 cm) or big size (65 to 100 cm), horizontal or vertical orientation, three-dimensional objects or two-dimensional pictures, visual interference by patterned paper versus white background. Desli and Giakoumi (2017) reported that the orientation of the TBEO was found to have an impact on the estimation ability: horizontal objects were estimated significantly better than vertically oriented objects. The other task characteristics had no significantly different influence on the students' length estimation ability. The results of Jones et al. (2012) differed from those of the study in primary school: Whereas no differences in the estimation accuracy could be observed between horizontally and vertically presented items, the dimensionality of the TBEO had an impact: it was easier to estimate the length of one side of a real three-dimensional cube in centimeter than it was on a 1:1-scaled picture of the cube.

Although Desli and Giakoumi (2017) varied the same task characteristics as Jones et al. (2012) did, they used different tasks, adapted to the age of the children they worked with. Another difference can be assumed in the cultural background of the students: Whereas Desli and Giakoumi (2017) investigated Greece 3rd and 5th graders, the initial inquiry of Jones et al. (2012) took place with U.S. middle school students. Due to these differences and the contradictory results in both studies, no clear statements concerning the characteristics of the TBEO to the accuracy of estimates can be made.

Concerning the *influence of the unit*, most British and American studies reported better results in estimating lengths in customary units than in metric units (e.g., Swan, & Jones, 1980; Jones et al., 2012). Since this is a special case concerning countries using non-metric units in their everyday life, we will not report all of them here in detail, but take a closer look to standard versus nonstandard units. Although overall the studies differ in the type of units they have included only few compared standard and nonstandard units. Desli and Giakoumi (2017) also varied the units systematically: as standard units, centimeters were used, as nonstandard units they presented pencils, straws, and paperclips. As Desli and Giakoumi (2017) reported, children's estimates were significantly more accurate (within a deviation range of 30 % of the real value) when being asked to use nonstandard units. Jones et al. (2012) also found overall differences between metric, customary and 'new' nonstandard units in their study with middle school students: estimates in metric units were the poorest in accuracy, which is probably due to the fact that these were less common in everyday life. Nevertheless, no clear pattern occurred in their post hoc analyses: some tasks were estimated better in one unit, and others were estimated better in one of the other units.

Concerning the *variation of other conditions* in lengths estimation studies, some characteristics of the TBEO, the unit and of benchmarks can be extracted from the procedure reported in the papers, but were not varied systematically: The TBEO can be present or absent, three-dimensional or two-dimensional, in real size or scaled, can be touched or not etc. Most of these conditions follow from the underlying definition of estimation or the use of the terms 'unit iteration' and 'benchmark'.

Forrester and colleagues investigated the role of 'context' in estimation tasks (Forrester et al., 1990; Forrester, & Shire, 1994; Pike, & Forrester 1997). Pike and Forrester (1997) compared classical textbook tasks – lines should be mentally measured by shorter lines as units – with contextually embedded narrative tasks: placing a number of ladybirds on a twig. Children performed better in textbook tasks, although the results missed the 0.5 significance level. Forrester et al. (1990) and Forrester and Shire (1994) also investigated the role of 'context' on primary children's estimates. But in every study their definition of 'context' differed: In 1990, length estimation tasks in steps, jumps and 'lying downs' were seen as real-world context in contrast to tasks asking for the estimation of the area of a triangle and a rectangle, seen as school situation context. In 1994, the authors investigated how former estimates and task conditions during the study – called 'context' – influenced later estimates, length estimation being a negligible part of it. All in all, we have no confirmed knowledge about the impact of 'context' on length estimation abilities.

Although more effort has been undertaken to understand measurement estimation abilities, it still remains unclear, which characteristics of the tasks are connected to which use of strategy, to which accuracy, and to which concept of length and measurement estimation at all. As already stated by other researchers (Joram et al., 1998; Sowder, 1992) a structured framework for the investigation of estimation abilities is still needed.

4 The Diversity of Estimation Tasks

As a first step to a theoretical model, possibly involved objects that can be part of a length estimation task and their characteristics were identified and are described here in detail. In a second step, these objects and their characteristics were combined in a tree diagram. For each path of this diagram, it is checked and discussed whether it leads to a meaningful estimation task or whether by fulfilling several features it becomes redundant and can be removed.

A complete estimation task needs naming the object and the attribute that should be estimated. Besides this typical estimation task, already formulated by Bright (1976), a basically different form of estimation task is formulated: the construction of an object of a given length (Fig. 1 on the left). In a paper-and-pencil test, this construction can be to draw a line. In other study designs, the subjects may be asked to cut off a piece of tape or rope with a defined length or similar actions.

Besides the requirement of a TBEO, an attribute and the question of construction or not, which are seen as basic characteristics of a measurement estimation task, other aspects can be supplemented, but are in most cases not absolutely necessary: marking a unit for estimation and indicating a possible benchmark for supporting the estimation process are supposed to be relevant. Just in one case, the unit has to be given: If the TBEO should be constructed, then the length of the TBEO, including the unit, has to be part of the estimation task.

For the further analysis of the characteristics, the following questions were discussed for each object: Is it given? Is it visible? Is it present in real size? Is it touchable? All questions were posed to each of the objects possibly being part of an estimation task.

4.1 The To-be-estimated-object (TBEO)

In a specific estimation situation, these three objects – the TBEO, the unit, and a possible benchmark – can be given in different configurations. As Bright (1976) already differentiated, the TBEO can be physically present (as a real object or as a real sized picture) or absent. In the first case, the TBEO is visible allowing different estimation strategies, which – in the case of lengths – is relevant, because estimation strategies for length are based on (mental) visual comparison. In the latter case the TBEO is not visible and has to be imagined (Fig. 1). In addition to this basic distinction some more characteristics of the TBEO should be taken into account: When the TBEO is physically present, one can further distinguish between touchable objects and not touchable objects. A third possibility is a TBEO that is represented by a scaled picture, so it is visible, but not in real size. In the last case, it does not make any difference, if the TBEO on that picture is touchable or not. Its real size is not touchable (Fig. 1).

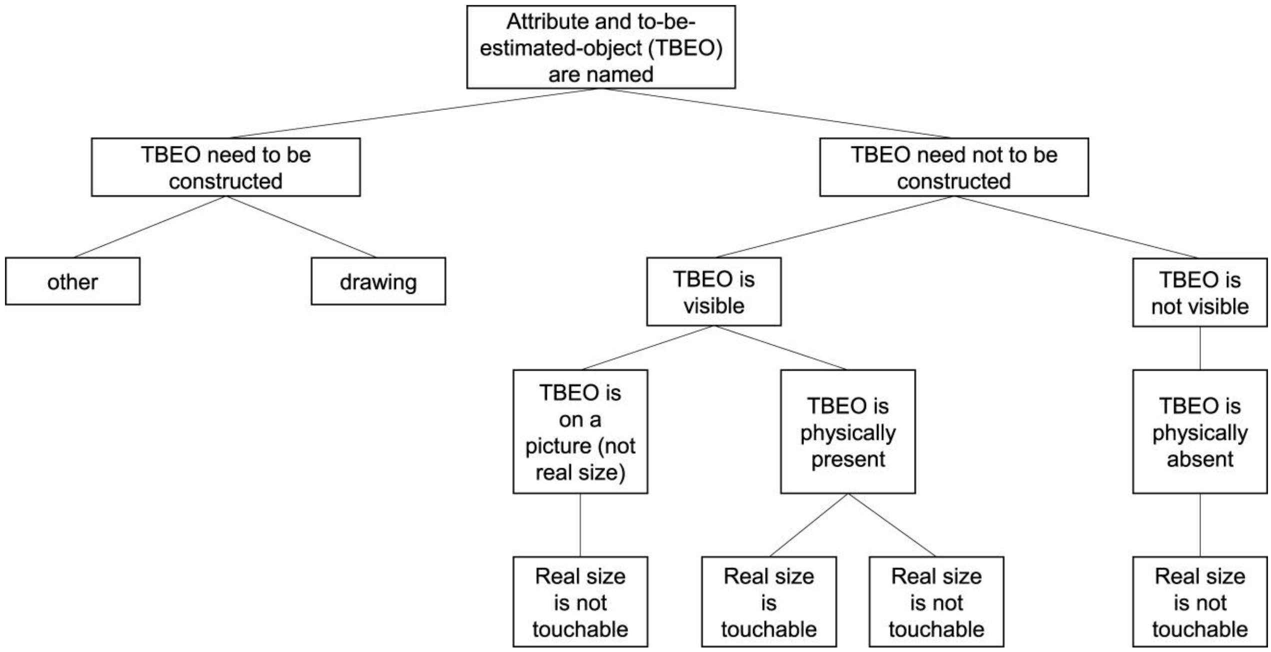

Fig. 1: Characteristics of the to-be-estimated-object (TBEO).

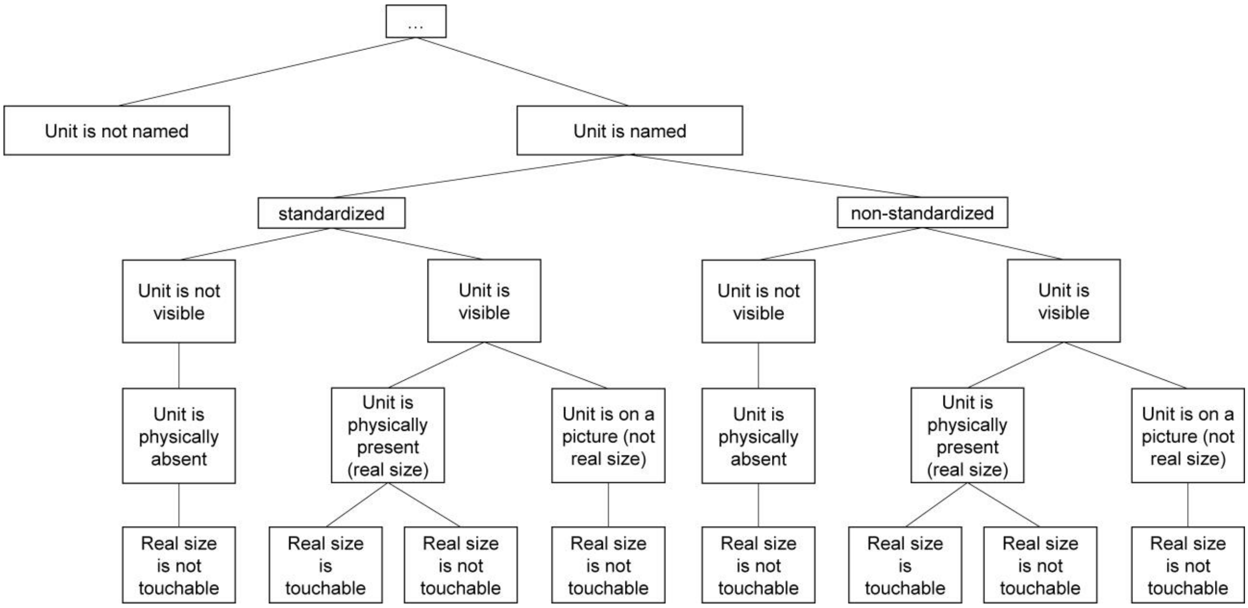

Fig. 2: Characteristics of the unit.

In constructing tasks, a special feature in the interpretation of item characteristics occurs, because they are changing during the construction process. At the beginning of the drawing process, the line is not physically present (and therefore not visible and not touchable), while after the drawing process it is physically present and therefore visible and touchable as well. This allows reflections about the estimation result in other ways than tasks in which the TBEO remains physically absent the whole time.

4.2 The Unit

The same characteristics are proposed for the unit (Fig. 2). By 'unit is visible' is meant that a representative of this unit, for example a line of one centimeter, is physically present. By 'unit is not visible' is meant that the unit is named, but there is no representation of it. So, the unit is absent as well. Like the TBEO, a unit that is physically present can be touchable or not. In addition, it is possible that the unit is not even named. Then, the estimator has to choose a proper unit.

Besides the pre- or absence of the unit, one other characteristic of units is important. As measurement can be done with standard units (e.g., metric units as mm, cm, m, km, or customary units as in, ft, yd, mi) and nonstandard units, for example with body measures or stripes, also affordances of estimation tasks can be differentiated in this way:

1) "How long is this line [picture]? This line is ____ cm long."

2) "How many of these stripes [picture] are as long as the longer side of the poster on the blackboard? ___ stripes are as long as the longer side of the poster on the blackboard."

In example (1), the unit is physically absent and therefore neither visible nor touchable. The unit – cm – is a standard one. In example (2), the unit is physically present, visible and touchable, because there is a representation of a special length printed in the testing booklet. The size of the TBEO should be estimated in stripes, which is a nonstandard unit.

Both standard and nonstandard units can also be used in drawing tasks:

3) "Draw a line which is 5 cm long."

4) "Draw a line which is as long as two of the stripes on the left [picture]."

Since in the first drawing task (3) standard units are used, the differentiation between the unit and the TBEO is clear, while there is no benchmark named: The TBEO is represented by the line and should be constructed. The standard unit 'cm' is not physically present.

In the second drawing task (4), the nonstandard unit 'stripe' is physically present and can be multiplied mentally to determine the length of the stripe to be drawn. Again, the TBEO is represented by the line.

But if the given nonstandard unit hasn't to be multiplied, the difference between TBEO and unit is difficult in drawing tasks, because every measurement process is a comparison and therefore a bidirectional relation (see Fig 3)

Fig. 3: Measurement as a bidirectional relationship: (3a) Is the pencil as long as 2 stripes or is the stripe as long as 0.5 pencils? (3b) Is the pencil as long as the stripe or is the stripe as long as the pencil?

This is illustrated by the following task:

5) "Draw a line which is as long as a paperclip."

There are two ways of looking at the TBEO and the unit: On the one hand, analogous to the examples above, the line could be seen as the TBEO and the paperclip as the unit. On the other hand, the paperclip could be seen as the TBEO and the line as the way to express the estimation result. Then, the line is seen as a standard-representation of a length, which is used instead of a measure. This could be seen as the construction of a unit (which is as long as the TBEO, so the measured value for the paperclip is '1 line').

Looking at linguistic similarities between the questions may help to solve this problem. The underlying structure is "do something which is like this measure". In tasks (3) and (4), the measures are 5 cm and 2 stripes. In accordance to this, the paperclip in task (5) cannot be seen as the TBEO but has to be the measure (1 paperclip). Therefore, paperclip is defined here as the unit, which may also be used with other parameters.

The third case, how a unit can be named, is as a scaled picture. The unit is then visible, but not in real size, and therefore the real size is not touchable. Real objects, which should be used as units, can then be represented by a photo so the need for resizing it is obvious. Standard units that are represented by lines cannot be represented as a picture, because the change of the size on a picture would automatically lead to a different unit. Just if the unit is represented by a real object, the resize is meaningful.

4.3 The Benchmark

As already mentioned by Bright (1976), a third object can be part of an estimation task in addition to the TBEO and the unit. This object is intended to be used as a benchmark (for a direct or indirect mental comparison). As described above, knowing a benchmark means knowing the measure of an object and having a mental image of its real size. While using an object as a benchmark both aspects are required. So, it can also be helpful to get the information about one aspect in the estimation task itself. The following examples give different suggestions:

6) "Imagine the longer side of a credit card. It is 8 cm long. How long is this sheet of paper? This sheet of paper is ___ cm long."

7) "On the left, you see a picture of a credit card in real size. How long is this sheet of paper? This sheet of paper is ___ cm long."

8) "On the left, you see a picture of a credit card in real size. It is 8 cm long. How long is this sheet of paper? This sheet of paper is ___ cm long."

In solving task (6), the credit card is intended to be used as a benchmark. The real size is not visible, but if the estimator can imagine its real size and combine this imagination with the measure named, they should be able to use it in the intended way. Question (7) includes a picture of a real-sized object, but the measure is not named and has to be imagined by the estimator. Task (8) includes both, a real-size-picture and the measure, so the object can be used as a benchmark even if the estimator had not mentally represented it as a benchmark yet.

In general, the object that can act as a benchmark can be characterized by the same features as the TBEO. In question (6), the benchmark is not physically present and therefore not visible or touchable. Questions (7) and (8) include a benchmark that is 'physically present', which means it is printed in real size in the test-booklet. Therefore, it is visible and touchable.

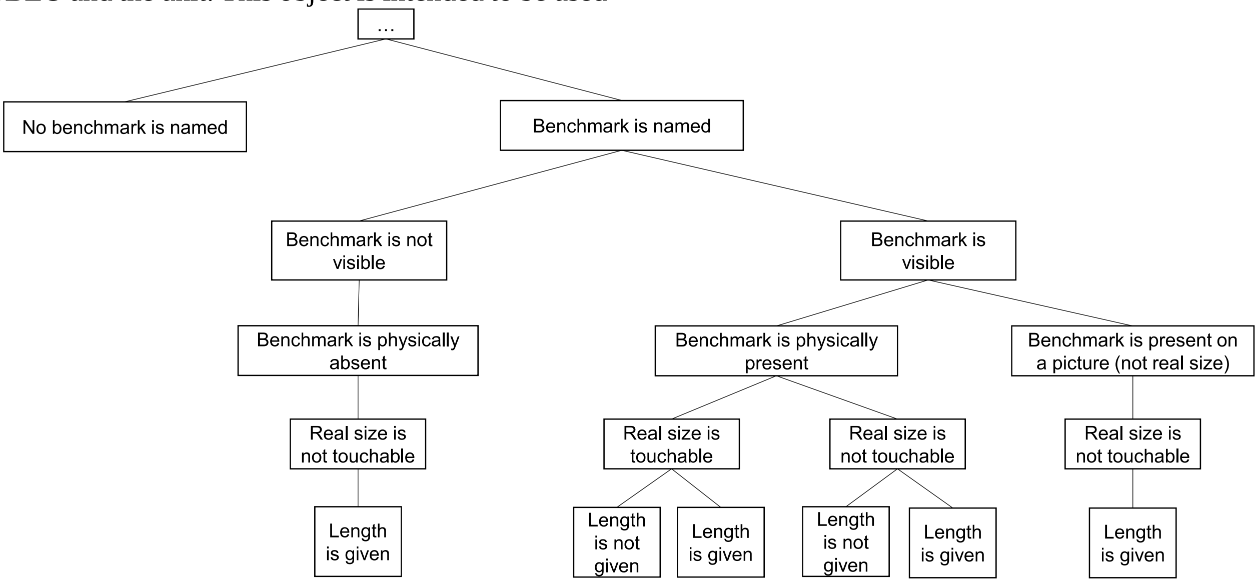

Fig. 4: Characteristics of benchmarks.

If the benchmark is not physically present and therefore not visible, or just represented by a scaled picture and therefore not visible in real size, the specification of its size is mandatory. Otherwise, no additional information that might be helpful in the estimation process would be given. If the benchmark is visible in real size, then a specification of its size may be helpful, but is not necessary due to the additional visual information provided.

Including the case that no benchmark is named, seven types for benchmarks can be distinguished (Fig. 4).

If the task mentions a benchmark, further assumptions are needed to distinguish it from a unit. For example, a stripe can serve as a benchmark:

9) "This stripe [picture] is 2 cm long. How long is the paintbrush at the blackboard? The paintbrush at the blackboard is ___ cm long. "

Seeing an object or a stripe as the unit, the parameter of this unit is 1 and no further steps of calculation are needed to get the answer of the iteration process. If the parameter is different from 1, it has to be multiplied with the number of iterations. Following this argumentation, the stripe in question (9) is seen as a benchmark. If the stripe had the length of 1 cm, it would have been seen as the unit.

5 Combination of all Characteristics – Building a Complex Model

Following the deliberations above, it becomes obvious that the eight types of estimation situation proposed by Bright (1976) are too broad categories to represent the variety of estimation tasks. If only three possibilities for included objects will be distinguished, the unit at least being standard or nonstandard and the implication of pictures and construction-tasks will be taken into account the model should be extended and revised. Furthermore, a structure of all important characteristics is needed to make a reasonable choice of tasks for a paper-and-pencil-test in length estimation or designing an intervention study etc.

5.1 Combining TBEO, Unit, and Benchmark in a Hierarchical Way

The easiest way to get a complete model that represents all different types of estimations is to combine all characteristics (Fig. 5).

The model is presented as a tree diagram. This kind of representation allows for analyzing the characteristics step by step. Following the paths helps to classify the tasks given or chosen.

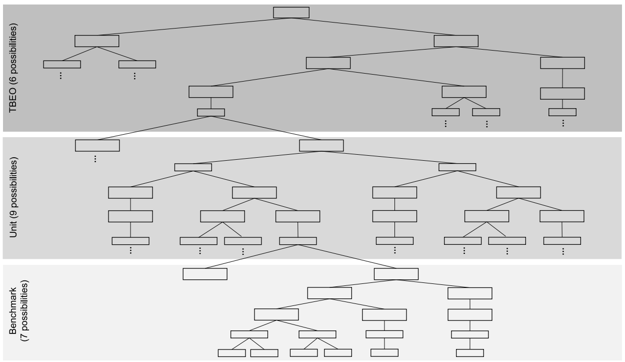

Fig. 5: Structure of the general model with the levels TBEO, unit, and benchmark and their characteristics.

The tree diagram as a whole consists of three levels: the TBEO with its characteristics, the unit with its characteristics, and the different types of benchmarks (Fig. 5). The TBEO is placed on top, at the first level, because without a TBEO there is no estimation task. The unit is put at the second level before the benchmark. Every estimate consists of a parameter of any kind of unit, and it is more likely that a unit is given in an estimation task than any kind of benchmark.

For any case of the tree, details about the visibility, the physical presence, the real size and, if needed, some specific characteristics can be explained and clarified. Theoretically, 378 estimation tasks are possible in this way. Nevertheless, some characteristics do not fit with others in some cases, so not all characteristics were included in all paths. For this reason, the number of possible estimation tasks is also reduced, as explained below.

5.2 Reduction of Possibilities by the Combination of Levels

In the complete model, most paths are similar to the explanations in chapter 4. However, the complete model is not just a combination of these possibilities, because some characteristics do not make sense in combination with others.

For each resulting path, an evaluation of its meaningfulness has to be done. To some extent, the exclusion or inclusion of any path depends on the aim for which the resulting tasks should be used. In order to develop a paper-and-pencil-test, it makes sense not to include for example both, a benchmark and the unit, to not over-determine the task. In school, for example to improve benchmark knowledge, maybe this is exactly the way to do it. The following descriptions of ex- or including different paths are based on the demand of the test development.

First of all, the combination of a to-be-constructed TBEO with the characteristics of the unit should be taken into account. If the TBEO should be drawn or otherwise constructed, the specification of a size is mandatory. A size specification always contains a unit. Therefore, the path 'unit is not named' is not possible for all construction tasks. The number of possible estimation tasks is therefore reduced by 14.

Different paths also occur as a consequence of previous characteristics and the demand for not to give too many details in the tasks. Therefore, at the level of benchmarks, the characteristics have to be adapted to the previous characteristics of the unit.

If the unit is not named, the length of a given benchmark must not be named to ensure that there is no indirect hint on the unit itself. If the length of the benchmark is given, it also includes a unit, so this would be contradictory to the condition 'unit is not named' at the previous level of the unit. The resulting paths of this case are presented in Figure 6.

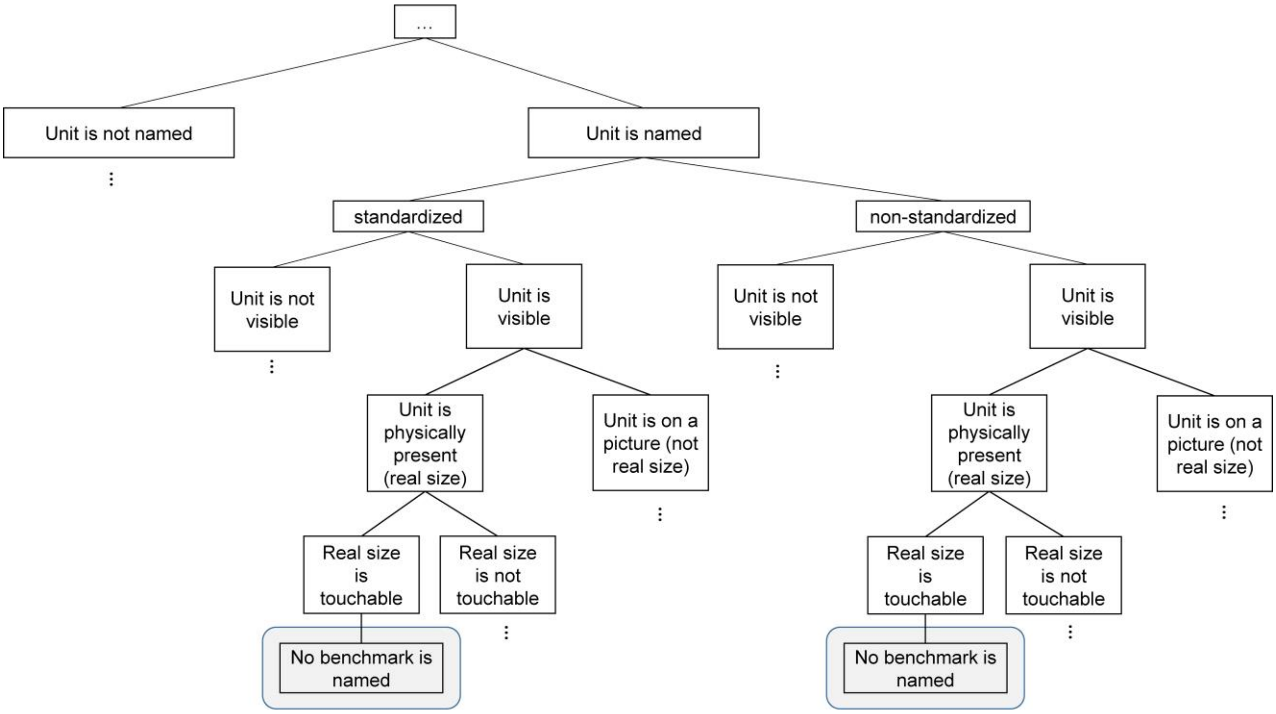

Fig. 6: Touchable units reduce the need for a benchmark.

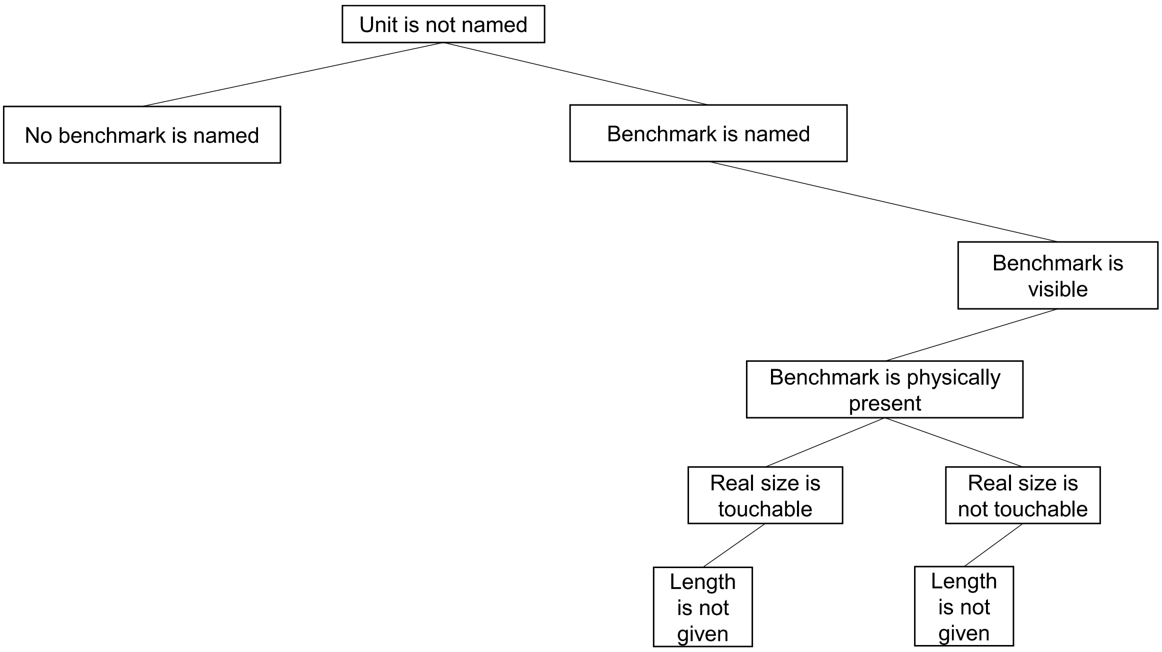

Fig. 7: Remaining possible benchmark situations if the unit is not named.

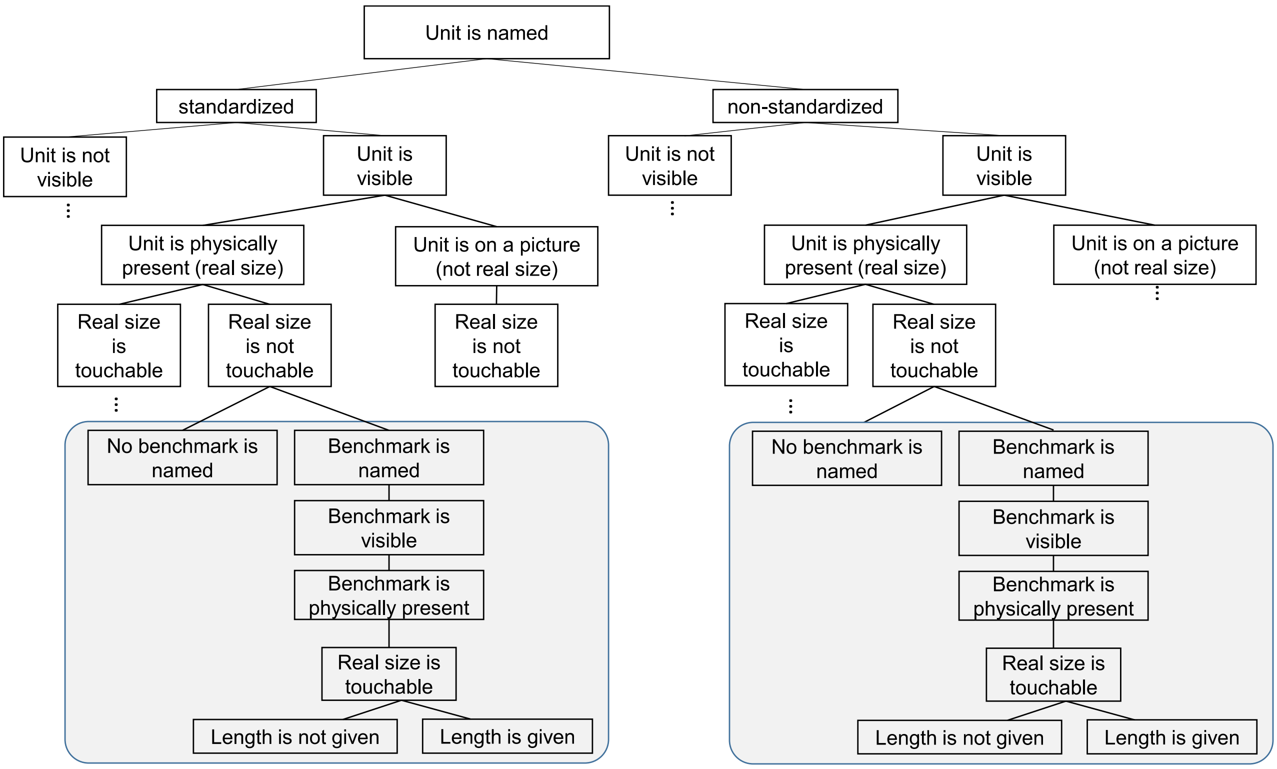

Fig. 8: Visible but not touchable units reduce the possibilities for benchmarks.

The number of estimation tasks is reduced by 16 (4 for each task in which the TBEO do not need to be constructed).

If the unit is given and touchable, there is no need for a benchmark at all (Fig. 7). Instead of using a benchmark and doing additional computations to get the result within the given unit one can use the unit itself for a direct estimation. So, there is no need for a benchmark, even independent of the question, if a standard or nonstandard unit is used. The number of estimation tasks is reduced by 72 (6 (benchmark) times 2 (unit) times 6 (TBEO)).

If the unit is given and visible in real size, but not touchable, there could be a given benchmark. However, in this case it is not enough to just name or show the benchmark. This would not be a simplification to just using the visible unit without further computations. To reach a simplification through more information, the given benchmark should be touchable (Fig. 8). The number of tasks is reduced by 48 (4 (benchmark) times 2 (unit) times 6 (TBEO)).

If the unit is given, but not visible in real size, any information about possible benchmarks could be helpful. Therefore, all seven characteristics for benchmarks provided in chapter 4 are possible (see fig. 4).

5.4 The Resulting Model

With regard to the previous deliberations, a complex model with some quite different paths could be created. In total, 228 remaining possible and meaningful estimation situations could be theoretically differentiated.

Although these many possible length estimation tasks could be theoretically meaningful, there has to be a choice of a subset of tasks for any purpose. However, the model itself helps to clarify which subset was chosen or is to be chosen.

6 Conclusion

Both the different estimation tests in cognitive psychology and mathematics education research and the theoretical development of this framework indicate that many different measurement estimation situations are used. As stated above, a theoretically based framework of possible estimation situations is needed for research in measurement estimation.

We considered three objects that will probably be part of any estimation task: a TBEO, a unit, and a benchmark. (While discussing each object with its characteristics, some possibilities could be removed as not being meaningful or even as logically not possible (section 4). When combining the three objects to a tree-diagram, further paths became overdetermined or even again not meaningful (section 5). Although many possibilities could be removed from the tree-diagram, the revised model still consists of 228 possible tasks.

The model is suggested to help researchers in at least two different ways: On the one hand, it may help during the planning of investigations, because it can be clarified with respect to the characteristic of the TBEO, the unit, and the benchmark which subset of tasks will be chosen for a special purpose. If the model is used with a special purpose, there may be more possibilities that can be removed or neglected, because of different reasons. Some possibilities may not be practical, for example in a whole-group paper-pencil-test, some may not be suitable for the subjects you are interested in, etc. On the other hand, the model may serve as an analytical tool for the comparison – or even for the clarification of the incomparability – of different studies.

Finally, a special case for touchable TBEO and touchable units or benchmarks should be discussed. If both, the TBEO and the unit (or benchmark), are touchable in their real size, concrete measurement processes are possible. If at least one object is movable, a direct measurement process can be carried out, if the objects are not movable, but touchable, an indirect concrete measurement process is possible through the help of another object (e.g., the fingers). Depending on the definition of estimation itself, 48 more task types will be removed – if you define estimation as strictly mental. Therefore, it also seems to be essential to always clarify the underlying definition carefully.

Even though the model already seems very complex and contains many characteristics, it is limited in some other ways. It still ignores possible variables that also may influence the estimation abilities. The most important characteristics concerning lengths that are not included yet can be seen in the size, the orientation and the dimensionality of the TBEO and the unit (and a possible benchmark). Although the results in other investigations did not show a clear tendency – Desli and Giakoumi (2017) did not get any differences between small (up to 30 cm) and big (65-100 cm) TBEOs –, it can be assumed that especially the ratio of the length of the TBEO and the unit that is asked for the estimate, matters. Scale may also make a difference: Very tiny lengths – up to some millimeters – may visually not be distinguishable anymore. Very big lengths, especially distances, will not be compared to stored images but estimated by other measures like knowledge about the time it takes to walk it.

Finally, yet importantly, the model so far is restricted to only one measurement area – length. The transferability of the framework to other measurement areas, especially to the visually perceivable geometric measures area and volume, can be assumed. However, this must be checked in detail. It also seems plausible that many the possibilities do not make sense for other measurement areas like time, speed or even mass/weight, because the attitude that has to be estimated is not visually perceivable. Especially the influence of visibility has to be discussed carefully in the latter cases as well as it has to be checked, if other characteristics become more important, for example the ratio of the mass and the volume in estimating the weight of an object.

The development of this framework and its limitations suggest that (length) measurement estimation is not a one-dimensional construct, unlike most studies seem – even unconsciously – to build on. Up to now, little effort has been undertaken to clarify the underlying construct systematically. The more insight we get into the construct of estimation abilities, the more we may stimulate necessary cognitive processes to foster and strengthen the (length measurement) estimation abilities of students.

Notes

[1] The paper deals with the systematization of objective task features as a contribution to the formulation of the complexity of possible estimation tasks. The extent to which this already captures and represents complexity is just as much an empirical question as the question of structuring possible estimation tasks with respect to subjective task features.

[2] In this text and the model developed here as well as in the literature mentioned, objects as TBEO mean representational objects. Besides, also for example the length of the distance between two objects could be estimated, but this is not applied in the model development here.

[3] As a reaction to estimation tasks, but not really an estimation strategy, *prior knowledge*, *visual impressions* and *indiosynctratic responses* are reported in the literature as well (Desli & Giakoumi, 2017; Hildreth 1983; Heid, 2018). Other researchers will call these non-strategic reactions as *inappropriate* or *no use of strategy* or *others* (Hildreth, 1980; Joram et al., 2005).

References

Axelrod, B. N., & Millis, S. R. (1994). Preliminary Standardization of the Cognitive Estimation Test. *Assessment, 1*(3), 269–274.

Brand, M., Fujiwara, E., Kalbe, E., Steingass, H.-P., Kessler, J., & Markowitsch, H. J. (2003). Cognitive Estimation and Affective Judgments in Alcoholic Korsakoff Patients. *Journal of Clinical and Experimental Neuropsychology, 25*(3), 324–334.

Bright, G. W. (1976). Estimation as Part of Learning to Measure. In D. Nelson & R. Reys (Eds.), *Measurement in school mathematics: 1976 yearbook* (pp. 87–104). Reston, VA: National Council of Teachers of Mathematics.

Bullard, S. E., Fein, D., Gleeson, M. K., Tischer, N., Mapou, R. L., & Kaplan, E. (2004). The Biber Cognitive Estimation Test. *Archieves of Clinical Neuropsychology, 19*(6), 835–846.

Corle, C.G. (1963). Estimates of quantity by elementary teachers and college juniors. *The Arithmetic Teacher, 10*(6), 347–353.

D'Aniello, G. E, Scarpina, F., Albani, G., Castelnuovo, G., & Mauro, A. (2015). Disentangling the relationship between cognitive estimation abilities and executive functions: a study on patients with Parkinson's disease. *Neurological Sciences, 36*(8), 1425–1429.

Della Sala, S., MacPherson, S. E., Phillips, L. H., Sacco, L., & Spinnler, H. (2003a). The role of semantic knowledge on the cognitive estimation task: Evidence from Alzheimer's disease and healthy adult aging. *Journal of Neurology, 251*(2), 156-164.

Della Sala, S.; MacPherson, S. E.; Phillips, L. H., Sacco, L., & Spinnler, H. (2003b). How many camels are there in Italy? Cognitive estimates standardised on the Italian population. *Neurological Sciences, 24*(1), 10–15.

Desli, D., & Giakoumi, M. (2017). Children's length estimation performance and strategies in standard and non-standard units of measurement. International *Journal of Research in Mathematics Education, 7*(3), 61–84.

Forrester, M. A., Latham, J., & Shire, B. (1990). Exploring Estimation in Young Primary School Children. *Educational Psychology, 10*(4), 283–300.

Forrester, M. A., & Shire, B. (1994). The Influence of Object Size, Dimension and Prior Context in Children's Estimation Abilities. *Educational Psychology, 14*(4), 451–465.

Goldstein, F. C., Green, J., Presley, R. M., O'Jile, J., Freeman, A., Watts, R., & Green, R. C. (1996). Cognitive Estimation in Patients with Alzheimer's Disease. Neuropsychiatry, *Neuropsychology, and Behavioral Neurology, 9*(1), 35–42.

Harel, B. T., Cillessen, A. H., Fein, D. A., Bullard, S. E., & Aviv, A. (2007). It takes nine days to iron a shirt: the development of cognitive estimation skills in school age children. *Child Neuropsychology, 13*(4), 309–318.

Heid, M. (2018). *Das Schätzen von Längen und Fassungsvermögen: Eine Interviewstudie zu Strategien mit Kindern im 4. Schuljahr* [Estimation of Length and Capacity: Strategies of 4th graders in an Interview Study]. Wiesbaden: Springer Spektrum (in German).

Heinze, A., Weiher, D. F., Huang, H.-M., & Ruwisch, S. (2018). Which estimation situations are relevant for a valid assessment of measurement estimation skills? 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.

Hildreth, D. J. (1980). *Estimation Strategy Uses in Length and Area Measurement Tasks by Fifth and Seventh Grade Students*. Dissertation at Ohio State University. Ann Arbor: University Microfilm International

Hildreth, D. J. (1983). The Use of Strategies in Estimating Measurements. *Arithmetic Teacher, 30*(5), 50–54.

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., Gardner, G. E., Taylor, A. R., Forrester, J. H., Andre, T., & (2012). Student's Accuracy of Measurement Estimation: Context, Units, and Logical Thinking. *School Science and Mathematics, 112*(3), 171–178.

Jones, M. G., Taylor, A. R., & Broadwell, B. (2009). Estimating Linear Size and Scale: Body Rulers. International *Journal of Science Education, 31*(11), 1495–1509.

Joram, E. (2003). Benchmarks as tools for developing measurement sense, in D. H. Clements (Ed.), *Learning and teaching measurement* (pp. 57–67), Reston (VA): NCTM.

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. *Review of Educational Research, 68*(4), 413–449.

Liss, M., Fein, D., Bullard, S., & Robins, D. (2000). Brief report: Cognitive estimation in individuals with pervasive developmental disorders. *Journal of autism and developmental disorders, 30*(6), 613–618.

MacPherson, S. E., Wagner, G. P., Murphy, P., Bozzali, M., Cipolotti, L., & Shallice, T. (2014). Bringing the Cognitive Estimation Task into the 21st Century: Normative Data on Two New Parallel Forms. *PLOS ONE. Advanced online publication*. [Retrieved from https://doi.org/10.1371/journal.pone.0092554]

Mendez, M. F., Doss, R., & Cherrier, M. (1998). Use of the Cognitive Estimation Test to discriminate Frontotemporal Demential from Alzheimer's disease. *Journal of Geriatric Psychiatry and Neurology, 11*, 2–6.

Mitchell, J. H., Hawkins, E. F., Stancavage, F. B., & Dossey, J. A. (1999). *Estimation skills, mathematics-in-context, and advanced skills in mathematics: Results from three studies of the National Assessment of Educational Progress 1996 mathematics assessment*. Washington, DC: National Center for Education Statistics.

O'Daffer, P. (1979). A Case and Techniques for Estimation: Estimation Experiences in Elementary School Mathematics – Essential, Not Extra! *The Arithmetic Teacher, 26*(6), 46–51.

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.

Shallice, T., & Evans, M. E. (1978). The Involvement of the Frontal Lobes in Cognitive Estimation. Cortex: a *Journal Devoted to the Study of the Nervous System and Behaviour, 14*(2), 294–303.

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. A project of the National Council of Teachers of Mathematics* (pp. 371-389). New York: Macmillan.

Swan, M. & Jones, O. E. (1980). Comparison of Student's Percepts of Distance, Weight, Height, Area, and Temperature. *Science Education, 64*(3), 297–307.

Weiher, D. F. & Ruwisch, S. (2018). Kognitives Schätzen aus Sicht der Mathematikdidaktik – Schätzen von visuell erfassbaren Größen und dazu erforderliche Fähigkeiten [Cognitive estimation from the perspective of mathematics education – the estimation of visually perceptable measures]. *mathematica didactica, 41*(1), 77–103 (in German).

Authors' Addresses

Dana Farina Weiher
Leuphana University Lueneburg
Institute of Mathematics and Mathematics Education
Universitaetsallee 1
21335 Lueneburg
Germany
weiher@uni.leuphana.de

Silke Ruwisch
Leuphana University Lueneburg
Institute of Mathematics and Mathematics Education
Universitaetsallee 1
21335 Lueneburg
Germany
ruwisch@uni.leuphana.de

Hsin-Mei E. Huang
University of Taipei
Department of Learning and Material Design
(with Master Program of Curriculum and Instruction)
1. AI-Gou W.Road
Taipei 10048
Taiwan R.O.C.
hhuang22@utaipei.edu.tw

Jessica Hoth
IPN – Leibniz Institute for Science and Mathematics Education
Olshausenstraße 62
24118 Kiel
Germany
hoth@ipn.uni-kiel.de

Aiso Heinze
IPN – Leibniz Institute for Science and Mathematics Education
Olshausenstraße 62
24118 Kiel
Germany
heinze@ipn.uni-kiel.de