JOURNAL OF RESEARCH IN SCIENCE, MATHEMATICS AND TECHNOLOGY EDUCATION
DOI: 10.31756/jrsmte.212

http://www.estej.com/

Framework for the Parallelized Development of Estimation Tasks for Length, Area, Capacity, and Volume in Primary School – A Pilot Study

Dana Farina Weiher

Leuphana University Lueneburg, Germany

Abstract: The purpose of this study is to present a framework for the development of parallelized estimation tasks for the visible measures length, area, capacity, and volume. To investigate if there are differences between the estimation types of task, a written estimation test for 3rd- and 4th-graders was developed. It includes eight different types of task for each measure. The percentage deviation of the estimated value from the real value (the measured size) of 137 students indicates that there are differences between the four measures as well as within the types of task that affect over- and underestimation and the estimation accuracy. Further research could address relations between the estimation of visible measures and the investigation of more characteristics in an estimation task, using a written estimation test that is based on this valid framework.

Keywords: *Estimation test; Estimation tasks; Measurement estimation; Parallelized items; Visible measures*

Introduction

Estimation in General

Estimation in general from a psychological point of view means answering a question whose exact answer is unknown. Therefore, it is necessary to use cognitive skills like, among others, developing an appropriate estimation strategy, reasoning, general knowledge, and executive functions (Brand et al., 2003; D'Aniello, Castelnuovo, & Scarpina, 2015; MacPherson et al., 2014). According to Winter (2003), who defines estimation from a mathematics education point of view, estimation can be described as a complex interaction between perceiving, remembering, correlating, rounding and calculating.

Psychological researchers focus on the parts of the brain involved during the estimation process. This focus results in a process model (Figure 1) that contains the working memory, the semantic long-term memory and the executive functions as parts for estimation (Brand et al. 2003; D'Aniello et al., 2015).

Mathematics education researchers mainly investigate the use of estimation strategies. They differentiate three kinds of estimation: number estimation, measurement estimation and computational estimation (Hogan & Brezinski, 2003; O'Daffer, 1979). Within measurement estimation, another distinction can be made. The estimating person can perceive the object's attributes (measures) such as length, area, capacity and volume with the eye. Therefore, these measures are characterized as visible measures. Other attributes of objects such as speed, time, weight and temperature are limited visible or not visible. Either visibility is expressed via other quantities, such as distance covered in a certain time (speed), or they must be made visible by performing an action (time), or they can only be perceived with other senses (weight, temperature).

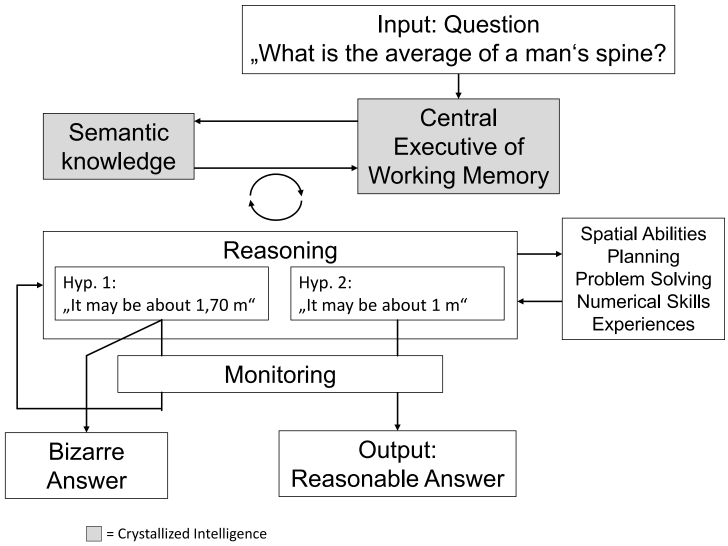

Figure 1. Model of Cognitive Estimation (D'Aniello et al., 2015).

Both psychological and mathematics education researchers use estimation tests for their investigations. Nevertheless, a valid theoretical base of how the tests were developed is often missing (Heinze, Weiher, Huang, & Ruwisch, 2018). In general, psychological tests mix up number and measurement estimation and, in addition, the different measures, without giving any further explanation (e.g. D'Aniello et al., 2015; MacPherson et al., 2014). Mathematics educations tests consider different kinds of estimation, but do not provide any references about the relation of different measures (e.g. Siegel, Goldsmith, & Madson, 1982).

Research Goal

This article focuses on the estimation of visible measures. Visible measures are length, area, capacity (within the meaning of liter and milliliter), and volume (within the meaning of cubic volume).

In current mathematical didactic research, there are no estimation tests based on a theoretically-based selection of different types of tasks, taking into account the possible differences of estimation of different sizes. For describing the estimation competence and the investigation of involved other competences like, for example, measuring competence or executive functions, it is necessary to use a valid estimation test. In addition, an estimation test for primary school aged children is missing.

The first aim of this article is to present a broad framework for the parallelized development of estimation tasks for length, area, capacity, and volume.

In addition, empirical results of the first test use are presented and discussed in order to obtain information on the suitability of the use in 3rd and 4th grade. More specifically, I sought answers for questions below:

- Are the types of task and measures suitable for 3rd- and 4th-graders?
- Which empirical differences between the characteristics of the tasks can be determined?

Theoretical Background

Measurement Estimation

Bright (1976) defines measurement estimation as "the process of arriving at a measurement without the aid of measuring tools. It's a mental process" (p.

89). Crucial in his explanation is the word *mental*. Already by doing a concrete measurement activity, the process is no longer seen as an estimation, but a measurement process.

This mental process is characterized by the comparison of the to-be-estimated-object (TBEO) with another object whose size and measure are known. These objects for comparison are named *benchmarks* (Joram, 2003). Most estimation strategies described for length, area, capacity, and volume are based on the process of comparison with benchmarks (Joram, Subrahmanyam, & Gelman, 1998; Siegel et al., 1982): Either the benchmark could be nearly the same size as the TBEO, or it has to be divided or multiplied. If this is not possible because the TBEO is too big or has a different shape, the estimator can simplify the estimation situation. Therefore, he could divide the TBEO to get an appropriate benchmark. Then, the person estimates the parts of the TBEO. Finally, the estimator merges the parts and their results to get the complete estimation result. This strategy is named decomposition/recomposition (Siegel et al., 1882). To bring the TBEO into a similar shape as the benchmark, the estimator can rearrange the TBEO mentally to simplify the comparison with the benchmark (Hildreth, 1983).

Additionally, a strategy for the estimation of area and volume exists: Length-Times-Width for area and Length-Times-Width-Times-Height for volume (Hildreth, 1983). These strategies are based on the formulas for rectangles and cubes. Therefore, the lengths of the sides of the rectangle respectively the edges of a cube are estimated and merged to get the result. Again, to estimate the length of the sides or edges, benchmark knowledge is required.

Even psychological studies indirectly refer to the use of benchmarks: They describe general knowledge to be relevant for accurate estimation results (e.g. Brand et al., 2003; D'Aniello et al., 2015). General knowledge could be seen as one part of the essential knowing about benchmarks (which means knowing the size of objects and be familiar with them).

Length, Area, Capacity, and Volume as Visible Measures

For the description of the relation between length, area, capacity, and volume two approaches can be used: first, as they are mathematics measures, the relation could be mathematically-derived. Second, because the estimation process of these measures is based, among others, on the understanding of the measurement process (Nührenbörger, 2004), an approach from the view of mathematics education can be useful.

The measures length, area, and volume (volume in the sense of a cube) are part of the same mathematical size system, with length as the base size. That means that all other measures can be derived from length. This becomes apparent by looking at the formulas for the computation of the size of an area or of a volume. For computing the size of an area, two lengths are multiplied, for computing the size of a volume, three lengths are multiplied. This fact is also reflected in the estimation strategies for area and volume. Capacity (in the sense of the content of a vessel) is not part of this size system, hence, it is more important in everyday life of young children than volume. Furthermore, it is part of the (German) primary school curriculum, so this meaning of volume should be part of a (German) visible measure estimation test.

In mathematics education research on estimation and measurement, some different reasons can be named for a joint investigation of estimation of length, area, capacity and volume.

First, all these measures are visible, and therefore some strategies that are already described for length can also be used to estimate the other measures (Figure 2). This is a result of the benchmark idea and the fact that length, area, capacity, and volume are visible measures. For all visible measures, the estimator can compare the TBEO with another object (the benchmark).

Second, the comprehension of the measurement process, which is one important aspect for the comprehension of the estimation process, includes similar aspects for the different measures. The most important aspect of measuring a size can be seen in iterating a unit of the same size. This idea is also possible for all visible measures. For length, e.g. lines as representatives of a unit can be laid end to end to one another. For area, squares as representatives for the unit can be used in the same manner, which creates a grid pattern. For volume, cubes can be used in the same way for a result without gaps and overlaps. Even if the process is not visible anymore after completion, this idea also works for capacity: A vessel can be used as a unit to measure the size of a bigger vessel. The number of pouring e.g. water from the smaller in the bigger vessel determines the size indication. After doing that, the number of pouring-activity is not visible anymore, but the idea of repeating a same sized unit is the same.

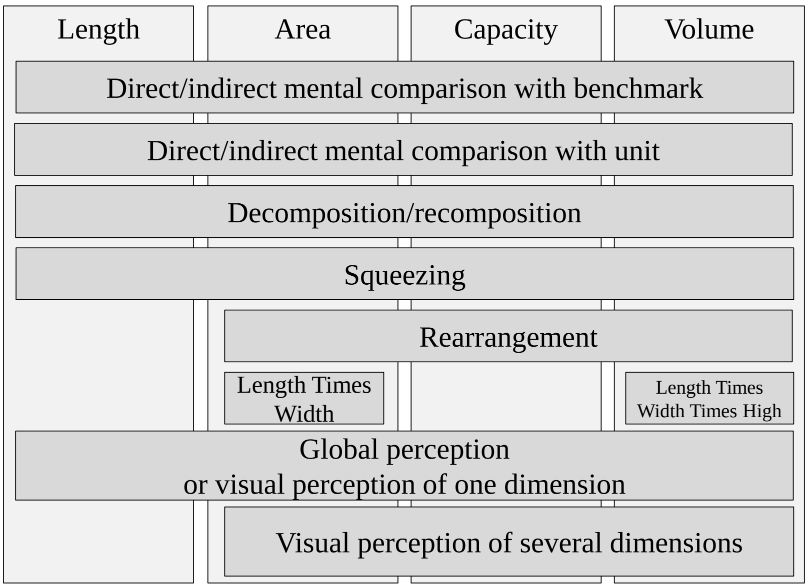

Figure 2. Estimation Strategies for the Measures Length, Area, Capacity, and Volume.

Third, Nührenbörger (2004) claims that the comprehension of the measurement process for length is not only similar to, but also fundamental for the comprehension of the measurement process from other measures, especially area and volume.

State of the Art: Different Types of Estimation Tasks

The first approach for structuring estimation tasks originates with Bright (1976). According to him, two objects are part of an estimation task: the TBEO, which measure should be estimated, and a unit. Both of these objects can be physically present or absent. Another type of task includes a measure to which an appropriate object should be found. For this task, a list with possible objects can be given (or not). Overall, Bright formulated eight types of estimation tasks (Figure 3).

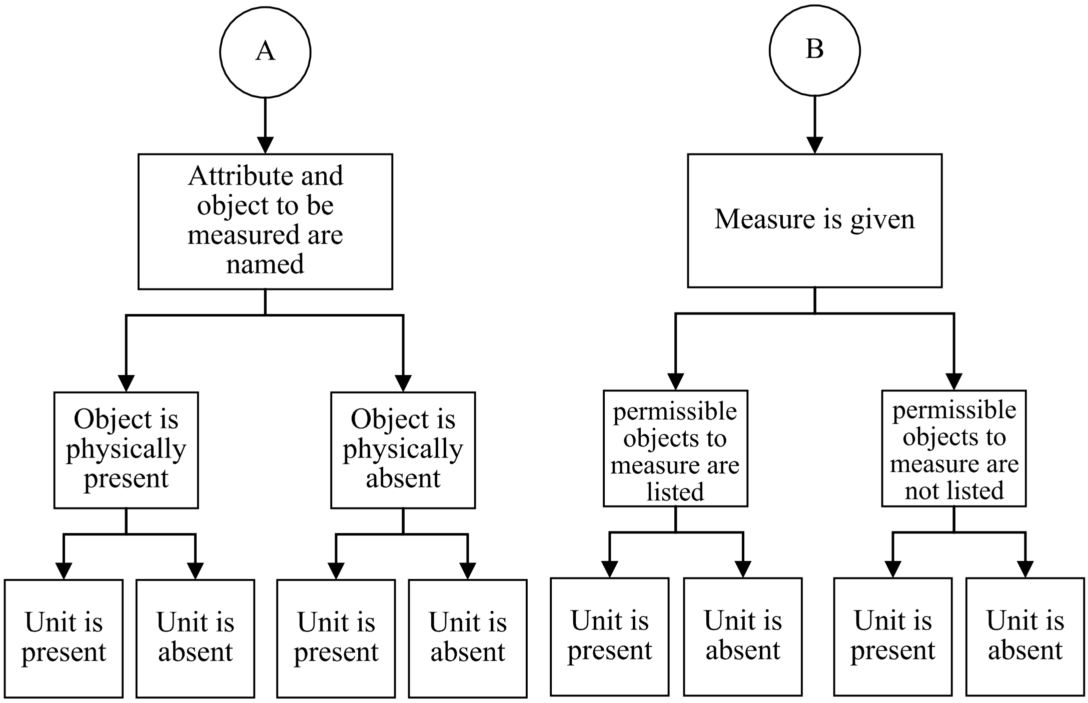

Figure 3. Eight Kinds of Estimation (Bright, 1976).

Heinze et al. (2018) used part A of Bright's model as initial position, to develop a broader framework for length estimation tasks. The aspect *physically present* distinguishes between *just visible* objects and *visible and touchable* objects. Besides that, *not physically present* automatically means *not visible*, but there is also the possibility to present the object with help of a picture (which means that it is *not physically present in real size*).

Another addition to Bright's model is a third object that can be part of an estimation task: the benchmark. This is an object of a given size that can be used as an object for a comparison. If the benchmark is given, the same characteristics as for the other objects are possible. It can be *physically present* or *not physically present*. If it is physically present, it can be *just visible* or *visible and touchable*.

The length estimation framework also includes construction tasks. These tasks require drawing a line with a length given. This entails that the TBEO is not visible at the beginning (because it is not constructed yet), but it is visible after the drawing process. Since these characteristics change during the working process, the distinction between visible and not visible is not appropriate for drawing tasks. Nevertheless, the other objects (unit and benchmark) can have the same characteristics as described above.

This framework is currently still restricted to length estimation. For creating parallelized items for the estimation of length, area, capacity, and volume, further development of the framework is needed.

A New Framework

Characteristics of Measurement Estimation Tasks

For developing an estimation test to investigate the visible measures length, area, capacity and volume, parallelized test items are desirable. This would be an improvement of existing tests (which mixed up both number and measurement estimation, and the different measures) and allows to get valid results. Furthermore, no valid estimation test exists for primary school aged students (who are the target group of this study) exists.

Parallelized items are characterized by equal task characteristics. They should also require the same or similar competences over the measures.

As described above, three objects can be part of all estimation tasks: the TBEO (which has to be named), a unit, and a benchmark. These objects can either be visible or not. For being visible, two possibilities exist: Either it is physically present (visible in real size), or it is shown on a picture (visible, but not in real size). If it is physically present, it can be touchable or not.

For developing parallelized items for all visible measures, three restrictions can be made from the very start.

First, the drawing tasks, which are included in the length estimation framework described above, do not fit the demand to address the same competences for every visible measure. It seems to be much easier to draw a line than to draw a cube, especially in a defined size. Because drawing is not a competence that is usually needed for estimation, it should not affect estimation tasks. By drawing lines, the impact could be left aside because all students should principally be able to draw a line. Therefore, drawing tasks could work well in estimating lengths, but not in estimating area, capacity, or volume. Consequently, the framework for all visible measures does not include drawing tasks.

Second, the framework does not include tasks with pictures of the TBEO, benchmark or unit. For length and area, there are no problems to perceive the real sizes and the relation between them from a picture, but for capacity and volume, there are. Due to the projection of three-dimensional objects on a two-dimensional surface, the estimator cannot perceive all real sizes or all real relations between them.

Third, if a benchmark is given, and should be usable as a benchmark, both the object and its size must be given. To use an object as a benchmark, the estimator has to have a clue about both the real size of the object and its measured size. If only one aspect is given, the benchmark may still be helpful (because it is more than nothing), but if both aspects are unknown to the estimator, the object given cannot be used as a benchmark. To ensure the difference between tasks with and without given benchmark, tasks that only include a picture of the benchmark or tasks that only name a size are not included in this framework.

Following these deliberations, the fundamental structure for possible estimation tasks for length, area, capacity, and volume, includes the characteristics shown in Figure 4.

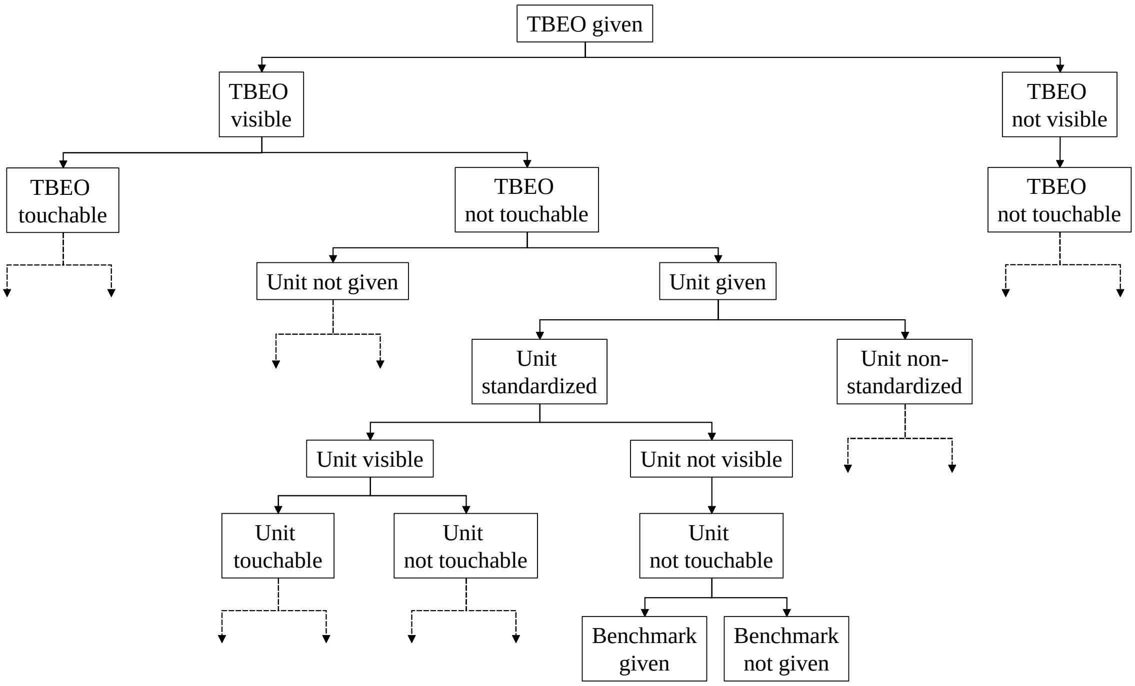

Figure 4. Possible Characteristics for Estimation Tasks.

By combination of the characteristics named above, overall 84 types of task result. Not all of them are appropriate for a written estimation test. The next chapter gives reasons for the exclusion of types of task.

Types of Task for the Parallelized Development of Estimation Tasks for Visible Measures

The most important reason to exclude a task from the estimation test is the possibility to do a measure. This applies to tasks that include two touchable objects (the TBEO and unit or TBEO and benchmark). Either the unit or the benchmark can be used to measure the size of the TBEO directly, or a third object, e.g. a finger, can be used to measure the unit or the benchmark first and the TBEO second. In both cases, this measuring process should be avoided. Consequently, all types of task with two touchable objects were excluded.

In some cases, a given characteristic entails another characteristic that should not be given. This is true if the unit should not be named, but the size of a benchmark is given, because the unit of the benchmark's size reveals the unit of the TBEO. To ensure the distinction between tasks with given a unit and tasks with no unit given, there are no benchmarks given if the unit is not named.

It can happen that more characteristics are given than needed. This is true if not only a visible unit, but also a benchmark is given. The benchmark is not necessary if the unit is visible, because for the estimation process, the unit could be used. So, if the unit is visible, a benchmark is redundant.

In tasks without a visible unit, it is theoretically possible to name an object that is intended to be a benchmark. Actually, both the size and the measure should be given (which means the object has to be visible) to make an object usable as a benchmark, but it could also be helpful to only know one aspect. This distinction is not made in the framework to avoid the situation that a student could not use the object in the intended way. Due to this unclear definition of "giving a benchmark" and, furthermore, the need of material and the number of items is high anyway, this type of task is excluded from the test.

Since we want to distinguish clearly between capacity and volume, the unit has to be named. If tasks with no given units are used for these measures it cannot be evaluated which concept the children are referring to. To ensure the parallelism between the four measures, this type of task is excluded not only for capacity and volume tasks, but also for length and area tasks.

Last but not least, the test is intended to be reasonable and understandable regarding the material. For length and area, touchable objects can be printed in the test booklet. For capacity and volume, printing is not possible because objects are three-dimensional. Consequently, real objects must be given to each student for all tasks that include touchable TBEOs, units, and benchmarks. This would increase the need of material (and therefore the costs and grasp) in an unacceptable way. Therefore, this kind of task was excluded for all measures.

Finally, eight of the 84 possible types of tasks were chosen (see Table 1).

			TBEO visible Not touchable	TBEO not visible
Unit standardized	Unit visible	Not touchable	Type 1	Type 2
	Unit not visible		Type 3	Type 4
Unit non-standardized	Unit visible	Not touchable	Type 5	Type 6
	Unit not visible		Type 7	Type 8

Figure 5. Eight Types of Task for Parallelized Items for Length, Area, Capacity, and Volume.

For this test, the TBEO can either be visible or not, but not touchable. There are two possibilities to fulfill these conditions: the TBEO may not be physically present or may physically be present, but will be shown at the blackboard. The pupils can see these objects, but they are not allowed to go to the front and touch it. The unit is always named. It is either visible or not, too, so for objects that represent the unit, the same characteristics apply as for the TBEO. The unit can be standardized or non-standardized.

Methodology for Testing the Suitability

Instrument

For investigating the suitability of estimation tasks for 3^{rd}- and 4^{th}-graders, four written estimation tests, one test for each measure, were developed. Each test includes eight types of task as described above. Three items represent each type of task. Overall, each test includes 24 items. The objects in these items should be familiar to students in 3^{rd} and 4^{th} grade. For each test, 45 minutes (one lesson) are provided.

The test includes materials, which are presented in the classroom. Minor changes in the array of materials are possible due to different furniture in the classrooms.

Participants

In this pilot-study, 137 children from three 3rd and three 4th grade classes were involved. The sampling was convenient. Not all children participated in all estimation tests; each class worked out two tests (which means two measures). Some students only filled in one test due to absence at one of the two dates of testing. This leads to different sample-sizes per measure (see Table 2).

Table 1.

Distribution of the Sample.

	Length	Area	Capacity	Volume
Class 3 girls	22	20	9	14
Class 3 boys	17	15	9	10
Class 4 girls	16	10	26	25
Class 4 boys	8	11	15	22

Data analysis

Missing values and outliers were counted to get information about the suitability of the measures for 3rd- and 4th-graders. Outliers were identified by using boxplots for each item (values below Q_1-3 IQR or above Q_3+3 IQR). They were deleted for further analysis to avoid distortion.

In order to investigate the suitability of the estimation tasks for 3rd- and 4th-graders, for each item the percentage deviation from the real value (P_r) of the TBEO was computed. Therefore, the real value (V_r) and the estimated value (V_e) is needed:

$$\frac{V_e - V_r}{V_r} \cdot 100\% = P_r. \qquad (1)$$

The deviation from the real value can be negative or positive, so “0%” would be the best result. A negative percentage deviation indicates an underestimation, whereas a positive percentage deviation means that there is an overestimation. To investigate the accuracy of an estimation, it is therefore necessary to use the modulus of the percentage deviation from the real size:

$$| P_r |. \qquad (2)$$

The Shapiro-Wilk test for testing normal distribution was used because of its high statistical power for small samples (Shapiro & Wilk, 1965). The results of the study are not normally distributed (Shapiro-Wilk test for length p = 0,018, area p = 0,091, capacity p = 0, volume p = 0).

Because the results are not normally distributed, percentiles of the percentage deviation were used to investigate over- and under-estimation and the accuracy of estimates for the different measures and different task characteristics.

Findings

Missing Values and Outliers

For each item, the missing values were counted to get information about the “attempting” to solve this item. Table 3 shows the number of missing values per measure.

Table 2.

Missing Values and Percentage per Measure.

Measure	Number of missing values (percentage)
Length	138 (9.12%)
Area	165 (12.27%)
Capacity	57 (4.03%)
Volume	634 (37.21%)

The high number of missing values concerning volume indicates that students have difficulties to estimate volume, or to understand what they have to do. This might be caused by not knowing the standardized units (cm^3, m^3, dm^3) that are used in the test booklet. Due to the observations during the test, and because it is not part of the curriculum of German primary school, it is possible that pupils do not really have a concept of this measure.

Regarding the lower numbers of missing values for length, area, and capacity, these measures seem to be less problematic. The difference between length and area on the one and capacity with less than half of missing values on the other side might be caused by the distribution of the sample. Two 3rd classes and one 4th class solved the length and area tasks, whereas one 3rd class and two 4th classes solved the capacity tasks (and the volume tasks).

Table 4 contains the number of outliers from the solved tasks. The number of extreme outliers is higher than the number of mild outliers for all measures. The highest amount of mild and extreme outliers are found in capacity tasks. Area estimation tasks have the lowest amount of mild outliers, while volume estimation tasks have the lowest amount of extreme outliers. This might be explained by the German curriculum again: because of the unknown standardized units or a missing concept of the measure, the competences concerning these measures might be quite similar. Consequently, a lower amount of outliers is the result.

Table 3.

Number of Outliers and Percentage per Measure.

Measure	Mild outliers (below Q_1-1,5 IQR or above Q_3+1,5 IQR)	Extreme outliers (below Q_1-3 IQR or above Q_3+3 IQR)
Length	38 (2.77%)	97 (7.06%)
Area	19 (1.62%)	62 (5.26%)
Capacity	60 (4.42%)	111 (8.17%)
Volume	36 (3.36%)	37 (3.46%)

Over- and Underestimation

To investigate the over- and underestimations, the percentage deviation from the real size was used. The arithmetic mean of the percentage deviation was computed for each student (and each measure). A negative arithmetic mean implies that this student tends towards under-estimation for this measure, whereas a positive arithmetic mean implies that this student tends towards over-estimation for this measure. For the investigation of the general over-and under-estimation of each measure, the arithmetic mean of all tasks of this measure for all students is used.

The arithmetic means of all students for length, capacity and volume are not normally distributed (verified using the Shapiro-Wilk-Test, length: p = 0.018, capacity and volume: p = 0.000), whereas for area, they are (p = 0.091). All curves are slightly positively skewed (skewness for length = 0.839, area = 0.741, capacity = 1.093, volume = 1.580). The skewness can be unattended because of the used scale is open to the right, but closed to the left (see discussion).

Table 5 shows the descriptive statistics of the percentage deviation from the real size for all measures. The arithmetic mean and the median indicate that length are rather over-estimated (positive arithmetic mean and median), while area, capacity and volume were rather under-estimated (negative arithmetic mean and median).

Table 4.

Descriptive Statistics for the Percentage Deviation from the Real Value per Measure.

Measure	*N*	*R*	*Min.*	*Max.*	*M*	*Med.*	*SD*
Length	63	142.20	-29.75	112.19	20.19	13.55	29.64
Area	55	145.31	-88.31	57.00	-28.79	-31.25	24.89
Capacity	59	127.14	-66.50	60.64	-25.85	-36.65	32.15
Volume	70	166.07	-93.00	73.07	-45.60	-58.08	39.02

By looking at the percentiles for each measure, this interpretation can be supported (Table 6).

Table 5.

Percentiles of the Arithmetic Mean of the Percentage Deviation from the Real Value per Measure.

	Length	Area	Capacity	Volume
N	63	55	58	69
Percentile 10	-14.23	-58.21	-58.65	-84.25
Percentile 20	-3.45	-48,52	-53.39	-74.07
Percentile 30	2.34	-42.01	-46.25	-69.5
Percentile 40	9,00	-34.84	-40.96	-65.11
Percentile 50	13.54	-31.25	-36.67	-58.75
Percentile 60	22.81	-26.80	-26.70	-47.42
Percentile 70	31.02	-19.85	-16.80	-35.60
Percentile 80	41.82	-6.67	-7.27	-24.5
Percentile 90	64.76	2.55	30.67	2.54

For length, all negative values (and probably a few positive values) are within the 30th percentile. This means, at most 30% of all students tend towards underestimation for the length of an object. For the other measures, the percentiles with all negative values are within the 90th percentile. This indicates that most students (nearly all) under-estimate these measures.

Over- and underestimation can be investigated also for the different types of task. In each case only one characteristic was investigated (independent of the other characteristics), because of the small number of items for each type of task and the small sample. Therefore, the tasks of each measure were divided up into two groups (e.g. TBEO visible or not). The arithmetic mean of the percentage deviation from the real size was computed for each group per student.

For visible TBEOs that length should be estimated, no consistent trend to over- or under-estimation could be indicated (Table 7). The transition from positive to negative arithmetic means is between the 50th and 60th percentile. That means that nearly the same number of students under- and over-estimate the length of a visible TBEO. Concerning the not visible TBEOs, the results show that students tend towards over-estimate this length. Less than 20% of the students under-estimate the length of a not visible TBEO.

In all other conditions, the measures of the TBEOs were in general under-estimated. More than 80% of the students under-estimate the size of a not visible TBEO in an area estimation task. Not visible TBEOs for capacity and volume were under-estimated in more than 90% of the cases. In the visible conditions, the number of students who generally under-estimate are slightly lower, but still the trend to under-estimation is obvious (area and volume more than 80%, capacity more than 60%).

Table 6.

Percentiles of the Arithmetic Mean of the Percentage Deviation from the Real Value in Estimation Tasks with Visible and Not Visible TBEO.

	Length TBEO visible	Length TBEO not visible	Area TBEO visible	Area TBEO not visible	Capacity TBEO visible	Capacity TBEO not visible	Volume TBEO visible	Volume TBEO not visible
N	63	63	55	55	59	59	69	70
Percentile 10	-24.96	-12.42	-77.96	-53.80	-48.50	-74.60	-86.83	-85.81
Percentile 20	-17.03	4.05	-65.49	-43.88	-39.91	-71.11	-73.67	-74.44
Percentile 30	-11.90	11.12	-58.80	-37.94	-35.17	-64.83	-70.78	-71.50
Percentile 40	-7.80	25.71	-52.28	-31.77	-26.50	-60.50	-63.50	-67.04
Percentile 50	-2.64	40.55	-45.33	-23.56	-15.81	-51.55	-57.67	-62.25
Percentile 60	2.53	51.35	-37.38	-17.90	-8.16	-48.42	-53.33	-57.10
Percentile 70	8.66	68.86	-26.62	-11.80	5.00	-36.13	-40.50	-43.65
Percentile 80	11.33	81.82	-19.77	-5.01	36.90	-31.50	-21.67	-31.47
Percentile 90	17.67	124.53	36.90	15.53	71.70	-11.33	39.41	-3.79

As the TBEO, the unit can be visible or not. The results (Table 8) indicate that there is no great difference between tasks with a visible unit and tasks with a not visible unit: In tasks with a visible unit, less than 50% of the students under-estimate the length of the TBEO. In tasks with a not visible unit, less than 40% of the students under-estimate the length of the TBEO.

Unlike to length estimation, it seems to make an obvious difference in the other measures if the unit is visible or not. More than 90% of the students under-estimate the TBEO in tasks with a unit that is not visible. For visible units, the number of under-estimating students is lower: for area, less than 40% of the students under-estimate the size of the TBEO if the unit is visible. For capacity and volume, less than 70% of the students under-estimate the size of the TBEO if the unit is visible.

Table 7.

Percentiles of the Arithmetic Mean of the Percentage Deviation from the Real Value in Estimation Tasks with Visible and Not Visible Unit.

	Length Unit visible	Length Unit not visible	Area Unit visible	Area Unit not visible	Capacity Unit visible	Capacity Unit not visible	Volume Unit visible	Volume Unit not visible
N	63	63	55	55	59	59	69	70
Percentile 10	-18.80	-21.29	-26.80	-83.09	-46.33	-77.00	-75.00	-95.49
Percentile 20	-13.20	-15.27	-13.17	-80.38	-41.08	-70.58	-57.83	-92.79
Percentile 30	-3.00	-6.17	-2.56	-77.00	-33.50	-67.40	-49.83	-90.59
Percentile 40	-0.16	0.99	2.80	-72.13	-23.00	-62.17	-38.67	-89.10
Percentile 50	4.33	8.75	15.00	-67.28	-16.40	-57.58	-32.25	-86.33
Percentile 60	19.82	21.30	21.33	-62.07	-7.27	-52.00	-19.83	-83.03
Percentile 70	36.78	34.41	35.04	-56.65	5.75	-40.10	4.00	-79.76
Percentile 80	49.34	43.61	66.76	-53.38	42.90	-28.27	32.33	-75.299
Percentile 90	97.7	55.53	102.29	-46.52	96.91	-14.00	93.33	-65.20

As a second characteristic, the estimation tasks include standardized or non-standardized units. Table 9 shows the percentiles of the arithmetic mean of the percentage deviation for the measures with the distinction of standardized and non-standardized units.

The percentiles indicate that for length, area, and capacity, more under-estimations were made in estimation tasks with standardized units. For length estimation tasks with standardized units, less than 50% of the students under-estimate the TBEO, so a tendency to over-estimation could be indicated. Nevertheless, the number of students that under-estimate the TBEO in tasks with non-standardized units, is lower (between 20% and 30%). For area and capacity, the number of students who under-estimate are higher: More than 90% under-estimate the area of the TBEO in tasks that include a standardized unit, and more than 80% under-estimate the capacity of the TBEO in tasks that include a standardized unit. The percentage of students who under-estimate the size of the TBEO when the unit is non-standardized is (slightly) lower: between 80% and 90% for area, and between 60% and 70% for capacity.

Table 8.

Percentiles of the Arithmetic Mean of the Percentage Deviation from the Real Value in Estimation Tasks with Standardized and Non-standardized Unit.

	Length Unit stand.	Length Unit non-stand.	Area Unit stand.	Area Unit non-stand.	Capacity Unit stand.	Capacity Unit non-stand.	Volume Unit stand.	Volume Unit non-stand.
N	63	61	52	55	57	59	36	70
Percentile 10	-27.64	-9.73	-90.83	-56.35	-75.38	-49.88	-96.90	-87.44
Percentile 20	-19.84	-2.26	-87.67	-48.74	-71.25	-40.58	-71.90	-78.18
Percentile 30	-12.28	6.50	-83.30	-34.20	-64.13	-35.73	-63.25	-75.61
Percentile 40	-2.34	13.41	-75.86	-28.14	-61.58	-30.25	-50.07	-70.36
Percentile 50	5.71	21.50	-68.14	-20.27	-52.92	-24.25	-36.40	-67.61
Percentile 60	11.09	37.03	-45.99	-13.57	-47.64	-6.92	7.77	-62.48
Percentile 70	17.89	47.13	-34.49	-6.27	-42.63	1.75	36.75	-57.17
Percentile 80	31.56	65.07	-19.58	-0.08	-26.50	22.18	81.41	-48.95
Percentile 90	62.43	105.50	-1.14	18.01	23.09	39.58	187.09	-28.55

In contrast, more than 90% of the students underestimate volume estimation tasks with non-standardized units, whereas less than 60% of the students underestimate volume estimation tasks with standardized units. These percentiles indicate that there is only a tendency to under-estimate the volume of an object if the unit is standardized, while there is an obvious trend to under-estimate the volume of an object if the unit is non-standardized.

Estimation Accuracy

For the investigation of the estimation accuracy, the modulus of the arithmetic mean of the percentage deviation from the real size was computed. Therefore, the modulus of the percentage deviation from the real size was computed for each task. The arithmetic mean of these values was computed for each child to evaluate the middle percentage deviation from the real value (the accuracy). Table 10 shows the descriptive statistics of the modulus of the percentage deviation from the real size per measure.

Length estimation tasks have in general the smallest percentage deviation from the real size (both the arithmetic mean and median are smaller than the others), while volume has the highest percentage deviation. The deviation for area and capacity is quite similar (both arithmetic mean and median).

Table 9.

Descriptive Statistics of the Modulus of the Percentage Deviation from the Real Value per Measure.

Measure	*N*	*R*	*Min.*	*Max.*	*M*	*Med.*	*SD*
Length	63	138.59	18.83	157.42	59.33	53.24	27.07
Area	55	68.45	48.83	117.28	70.04	68.80	13.47
Capacity	59	186.13	42	228.13	74.88	65.79	28.57
Volume	70	159.13	43.63	202.76	86.58	74.91	35.45

The standard deviation is the smallest for area (SD = 13.47) and the highest for volume (SD = 35.45), while the range is the highest for capacity (186.13). This indicates that student's estimates are nearly similar for area and quite different for volume and capacity estimation tasks. The standard deviations for length (SD = 27.07) and capacity (SD = 28.57) are quite similar. This indicates that the student's percentage deviations are similar within length and capacity.

The arithmetic means from the modulus of the percentage deviation from all students for all measures are not normally distributed (verified by using the Shapiro-Wilk Test, length: p = 0.001, area: p = 0.011, capacity and volume: p = 0.000). All curves are slightly positively skewed (skewness for length = 1.153, area = 0.959, capacity = 3.106, volume = 1.949). The skewness can be unattended because of the used scale is open to the right, but closed to the left (using the modulus of the percentage deviation might reinforce this effect).

For investigating the accuracy of the estimates per measure, Table 11 shows the percentiles of the modulus of the percentage deviation from the real size for each measure.

Table 10.

Percentiles of the Arithmetic Mean of the Modulus of the Percentage Deviation from the Real Value per Measure.

	Length	Area	Capacity	Volume
N	63	55	59	70
Percentile 10	31.14	52.96	54.20	59.03
Percentile 20	37.20	57.39	58.96	62.90
Percentile 30	42.27	60.36	60.05	66.85
Percentile 40	48.13	65.89	61.34	71.57
Percentile 50	53.23	68.80	65.79	74.91
Percentile 60	60.98	72.81	70.72	82.09
Percentile 70	70.06	77.63	76.25	88.53
Percentile 80	77.24	79.67	85.17	99.93
Percentile 90	101.45	87.60	114.30	131.57

These results indicate that area estimation tasks were estimated most accurate (90% of the students show a maximal deviation of 87.6% from the real value) followed by length (90% of the students have a maximal deviation of 101.45% from the real value). However, it is conspicuous that the best 10% of students estimate area with a maximal deviation of 52.96%, while in length estimation tasks, 50% of the students reached a similar deviation (53.23%).

The lowest estimation accuracy is shown for capacity (90% of the students have a maximal deviation of 114.3%) and volume (90% of the students have a maximal deviation of 131.57%). This is supported by the fact that only 10% of the students have a deviation of 54.2% (capacity) or 59.03% (volume). For length and area, the percentage of students with similar deviation is higher.

The accuracy of the estimates can also be investigated for each type of task. The percentiles indicate that for area and volume, it makes no noticeable difference if the TBEO is visible or not. The percentage deviation from the real value is nearly the same for both type of task (Table 12).

Table 11.

Percentiles of the Arithmetic Mean of the Modulus of the Percentage Deviation from the Real Value in Estimation Tasks with Visible and Not Visible TBEO.

	Length TBEO visible	Length TBEO not visible	Area TBEO visible	Area TBEO not visible	Capacity TBEO visible	Capacity TBEO not visible	Volume TBEO visible	Volume TBEO not visible
N	63	63	54	55	59	59	69	70
Percentile 10	21.41	30.96	51.94	50.59	44.64	58.89	53.33	59.08
Percentile 20	26.63	39.05	56.17	58.11	49.33	61.75	57.67	64.00
Percentile 30	30.60	46.00	60.52	59.42	50.60	66.33	64.33	69.37
Percentile 40	34.20	54.60	62.58	63.98	56.00	70.80	68.75	71.12
Percentile 50	36.82	58.90	67.81	68.57	56.67	73.67	73.67	75.95
Percentile 60	40.93	78.71	73.33	72.29	59.67	77.00	78.50	80.73
Percentile 70	47.81	101.98	77.85	77.33	66.17	78.82	86.58	85.72
Percentile 80	56.83	117.50	80.67	81.47	95.80	85.41	93.42	91.63
Percentile 90	64.78	152.15	92.95	90.43	128.00	95.87	132.83	131.63

In contrast to area and volume, for length and capacity there seem to be a difference. Most of the students have a deviation of 64.78 % or less if the TBEO is visible, whereas they have a deviation of up to 152.15% if the TBEO is not visible. This indicates that length estimation tasks with a visible TBEO were estimated more accurate than length estimation tasks with not visible TBEOs. For capacity, it is the other way round. If the TBEO is visible, 90% of the students show a deviation up to 128%. If the TBEO is not visible, 90% of the students have a deviation of 95.87%.

The estimation accuracy varied within the different characteristics for the unit (Table 13). For length estimation tasks, the deviation is lower if the unit is visible (90% of the students have a deviation of 114.43% or less) than if the unit is not visible (90% of the students have a maximum deviation of 118.77%). For area, capacity and volume, most of the students have a smaller percentage deviation in tasks with visible object. This trend reverse not until the 80th percentile (area, volume) and the 60th percentile (capacity).

Table 12.

Percentiles of the Arithmetic Mean of the Modulus of the Percentage Deviation from the Real Value in Estimation Tasks with Visible and Not Visible Unit.

	Length Unit visible	Length Unit not visible	Area Unit visible	Area Unit not visible	Capacity Unit visible	Capacity Unit not visible	Volume Unit visible	Volume Unit not visible
N	63	63	61	55	59	59	69	70
Percentile 10	27.37	23.28	39.60	56.73	49.13	49.82	42.60	68.82
Percentile 20	31.23	33.27	46.04	61.60	56.11	60.08	46.33	75.77
Percentile 30	33.34	40.29	52.30	67.01	59.44	64.73	49.83	84.33
Percentile 40	38.05	48.93	58.30	69.57	64.43	69.45	56.71	86.07
Percentile 50	44.75	56.00	63.00	75.45	69.25	72.42	61.00	88.44
Percentile 60	49.86	65.08	67.25	77.18	76.33	74.33	71.33	90.17
Percentile 70	61.85	75.62	73.39	79.18	83.57	75.58	81.25	91.45
Percentile 80	84.93	96.53	83.95	81.98	94.50	77.67	115.67	92.98
Percentile 90	114.43	118.77	91.25	85.74	111.50	82.10	178.36	96.84

The last characteristic to look for intensively is the difference between standardized or non-standardized unit (Table 14). The percentage deviation from the real size is higher for estimation tasks for area, capacity, and volume, if the unit is non-standardized. This is valid for all percentiles, but especially obvious for volume estimation tasks (90% of the students have a percentage deviation of 295% from the real size if the unit is standardized, whereas the deviation is only 87% if the unit is non-standardized). In capacity estimation tasks, this trend is also noticeable.

For length estimation tasks, the trend is not that noticeable. For the most accurate estimates, the unit is non-standardized, but with increasing inaccuracy, the tasks with standardized units were estimated better than tasks with non-standardized units.

Table 13.

Percentiles of the Arithmetic Mean of the Modulus of the Percentage Deviation from the Real Value in Estimation Tasks with Standardized and Non-standardized Unit.

	Length Unit stand.	Length Unit non-stand.	Area Unit stand.	Area Unit non-stand.	Capacity Unit stand.	Capacity Unit non-stand.	Volume Unit stand.	Volume Unit non-stand.
N	63	61	52	55	57	59	36	70
Percentile 10	29.74	23.10	67.25	38.99	55.64	45.58	67.63	54.72
Percentile 20	33.82	29.76	71.05	43.45	59.73	50.91	75.12	62.23
Percentile 30	38.47	38.98	74.58	49.65	64.30	53.58	87.05	64.44
Percentile 40	43.65	42.98	80.13	54.23	67.00	59.17	92.23	67.09
Percentile 50	47.25	49.71	83.75	62.44	71.00	61.18	99.17	71.03
Percentile 60	51.80	60.57	87.16	66.88	73.81	64.88	115.10	74.15
Percentile 70	57.03	71.86	88.65	70.07	79.65	68.73	173.31	76.52
Percentile 80	72.53	92.63	90.83	76.18	96.67	78.80	187.92	80.95
Percentile 90	108.53	112.80	96.19	88.60	133.84	97.42	295.48	87.44

Discussion

Estimation of length, area, capacity, and volume have different results concerning under- and overestimation and estimation accuracy. Under- or over-estimation of the measures in general might be caused by the limitations or inexperience by using a higher number range. For area, capacity and volume, the measure values were naturally higher because of more dimensions. A consequence for not feeling safe using higher numbers (for everyday-sized objects) might be choosing lower numbers as measured values which results in an under-estimation. The everyday-experience with length may compensate the insecurity with higher numbers. Students might feel safer using higher length-sizes than high numbers for measures they do not have much experience with. This conjecture is supported by the fact that the mean of the modulus of the percentage deviation is smaller for length than for the other measures (see below).

Higher accuracy for not visible TBEOs than for visible TBEOs in capacity estimation tasks might be explained with the higher number of under-estimations in these tasks. An under-estimation has a maximal deviation of 100%. When the TBEO is visible, over-estimations are more frequently. This result has an impact on the highest deviation (over-estimations are not percentage limited). The reverse argumentation can be used for length: Tasks with not visible TBEOs were in general over-estimated, therefore, the accuracy is lower. In addition, it is more difficult to estimate the size of a not visible object because more cognitive processes (like getting an appropriate image of the object and its size from the memory) are necessary.

In tasks with not visible units, the area, capacity or volume of an object was more often under-estimated than in tasks with visible units. This difference might be caused by mental benchmarks that are too small or because by the absence of any benchmark for these measures (and wild guessing causes estimated that are too low because of the higher number range, see above). If a unit is visible, the estimation is supported by a hint of the real magnitude, so it is easier to estimate. These results might cause a better accuracy in tasks with not visible units for area, capacity and volume.

Estimation tasks for capacity and area with non-standardized have a better accuracy, although these tasks were in general underestimated. For area, this might be explained with the unknown standardized units, but the units for capacity actually should be known. The accuracy in volume estimation tasks with standardized units is noticeable lower than in volume estimation tasks with non-standardized units. This might be explained on the one hand with the unknown standardized unit, on the other hand by the general over-estimation in volume estimation tasks with standardized units.

Because of the different results concerning over- and under-estimation and estimation accuracy, the parallelism between the tasks and the measures should be discussed. On the one hand, estimation of visible measures seems to require similar competences because the estimation strategies could be used for all measures. On the other hand, even these visible measures have differences like, e.g., the dimensionality. The results indicate that measurement estimation in general is not a valid construct and should be theoretically divided in length estimation, area estimation, capacity estimation and volume estimation.

The different types of task per measure cause different evaluations of the appropriateness for 3rd- and 4th-graders. The high amount of missing values in volume estimation indicates that these tasks are not appropriate for 3rd- and 4th-graders. This seems to be limited to tasks with standardized units: 70 from 71 students attempt tasks with non-standardized units, whereas only 36 students attempt tasks with standardized units. For area, this difference is not noticeable.

Conclusion

The most important conclusion is the fact, that the one and only estimation task does not exist. Different types of task result in different (strong) tendencies for over- and under-estimation and different estimation accuracies. This pilot study shows that an estimation test must include different types of task to get a valid result.

Furthermore, this pilot study indicates that 3rd- and 4th-graders are able to solve estimation tasks for length, area, capacity, and volume. Except tasks with standardized units for volume, they are able to deal with all types of task. However, because this might be caused by the German curriculum which does not involve standardized units for this measure in primary school, in general, these tasks might be usable in higher grades.

Further research is needed to investigate the relations of all characteristics among each other. Therefore, a bigger sample and even more items per type of task are needed to allow factor analysis for the eight types of task in this estimation test.

Finally, the question of scoring is not answered yet. Different studies in psychology and mathematics education research do not only use different tasks, but also different kinds of scoring. Consequently, the comparison of these studies' results is difficult (Clayton 1996). Further research is needed to improve and develop an appropriate scoring for (parallelized) items for all visible measures.

References

Bright, G. W. (1976). Estimation as Part of Learning to Measure. In D. Nelson (Ed.), *Measurement in School. Mathematics 1976 Yearbook* (pp. 87–104). Reston, VA: National Council of Teachers of Mathematics.

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. doi: http://dx.doi.org/10.1076/jcen.25.3.324.13802

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. doi: http://dx.doi.org/10.3389/fpsyg.2015.00608

Heinze, A., Weiher, D. F., Huang, H.-M., & Ruwisch, S. (2018). Which Estimation Situations are Relevant for a Valid Assessment of Measurement Estimation Skills? *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education (Vol. 1)*. Umeå, Sweden: PME.

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. doi: http://dx.doi.org/10.1207/S15327833MTL0504_02

Joram, E. (2003). Benchmarks as Tools for Developing Measurement Sense. In D. H. Clements & G. Bright (Eds.), *Learning and Teaching Measurement. 2003 Yearbook* (pp. 57-67). Reston, VA: National Council of Teachers of Mathematics.

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. doi: http://dx.doi.org/10.3102/00346543068004413

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 9(3): e92554. doi: http://dx.doi.org/10.1371/journal.pone.0092554

Nührenbörger, M. (2004). Children's Measurement Thinking in the Context of Length. In G. Törner, R. Bruder, A. Peter-Koop, N. Neill, H. G. Weigand, & B. Wollring (Eds.) *Developments in Mathematics Educations in German-speaking Countries. Selected Papers from the Annual Conference on Didactics of Mathematics. Ludwigsburg 2001.* (pp. 95-106).

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.

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Sample). *Biometrika, 52* (3/4), 591-611. doi: http://dx.doi.org/10.2307/2333709

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. doi: http://dx.doi.org/10.2307/748557

Winter, H. (2003). *Sachrechnen in der Grundschule. Problematik des Sachrechnens. Funktionen des Sachrechnens. Unterrichtsprojekte*. (6th ed.). Frankfurt am Main: Cornelsen Scriptor.

Corresponding Author Contact Information:

Author name: Dana Farina Weiher

University, Country: Leuphana University Lueneburg, Germany

Email: weiher@leuphana.de

Please Cite: Weiher, D. F. (2019). Framework for the Parallelized Development of Estimation Tasks for Length, Area, Capacity, and Volume in Primary School – A Pilot Study. *Journal of Research in Science, Mathematics and Technology Education, 2*(1), 9-28. doi: 10.31756/jrsmte.212

Received: October 19, 2018 ▪ Accepted: January 10, 2019