for-granted understandings of education. I simply want to suggest that our ordinary understanding of reviews is perfectly compatible with a modernist conception of education: a review sums up or synthesizes knowledge so that we can enhance our capacity to solve educational problems. Education is under our control, and the knowledge we aspire to is a kind of *techne*. But if education is about self-understanding modeled on *phronesis*, and if acquiring self-understanding is a hermeneutic process in which the "unfamiliar and the familiar, the learner and a particular tradition, the possible and the real, are all challenged and opened up" (Gallagher, 1992, p. 189) in a process of questioning, then we need the kinds of reviews that "put us in the way" of education and invite our participation (R. Peters cited in Gallagher, p. 185).

References

Carr, W. (1995). *For education: Towards critical educational theory.* Buckingham, England: Open University Press.

Dunne, J. (1993). *Back to the rough ground.* Notre Dame, IN: University of Notre Dame Press.

Gadamer, H.-G. (1981). *Reason in the age of science*, trans. F. G. Lawrence. Cambridge, MA: MIT Press.

Gadamer, H-G. (1989). *Truth and Method.* 2nd rev. ed., trans. J. Weinsheimer and D. G. Marshall. New York: Crossroad Publishing Company.

Gallagher, S. (1992). *Hermeneutics and education.* Albany, NY: SUNY Press.

Grondin, J. (1995). *Sources of hermeneutics.* Albany, NY: SUNY Press.

Strike, K., & Posner, G. (1983a). Epistemological problems in organizing social science knowledge for application, in S. Ward & L. Reed, Eds., *Knowledge structure and use.* Philadelphia: Temple University Press.

Strike, K., & Posner, G. (1983b). Types of syntheses and their criteria, in S. Ward & L. Reed, Eds., *Knowledge structure and use.* Philadelphia: Temple University Press.

Review of Educational Research
Winter 1998, Vol. 68, No. 4, pp. 413–449

Measurement Estimation: Learning to Map the Route From Number to Quantity and Back

Elana Joram
University of Northern Iowa

Kaveri Subrahmanyam
California State University, Los Angeles

Rochel Gelman
University of California, Los Angeles

Measurement estimation has been identified as a critical area for mathematical development, yet little is known about how estimators make their judgments and how competency in measurement estimation can be supported through instruction. We present a framework in which skilled estimators move fluently back and forth between written or verbal linear measurements and representations of their corresponding magnitudes on a mental number line. With reference to the proposed framework, a review of instructional efforts to teach measurement estimation suggests they are incomplete and are limited to producing associations between isolated measurements and their referents. Recommendations are discussed for instruction that would extend students' understanding beyond these discrete points to conceptualize measurement scales and systems.

Most adults estimate measurements every day of their lives. We judge how long it will take to get to work on time, read newspaper articles cautioning us to be wary of skin moles more than a quarter-inch wide, or determine that a bunch of grapes we select at the supermarket weighs approximately a pound–these varied activities all involve measurement estimation. Children, too, learn at an early age to talk about estimated quantities: for example, that she or he is a little shorter than a brother or sister or that a carton of milk can fill more than three glasses (National Council of Teachers of Mathematics, 1989).

Measurement estimation can be thought of as *physical measurement* without tools (Bright, 1976). Whereas the use of a measuring instrument such as a ruler

The preparation of this paper was supported by a Small Grant from the Spencer Foundation and by a Summer Research Fellowship from the University of Northern Iowa to the first author. We would like to thank Diane Thiessen and three anonymous reviewers for helpful comments on an earlier draft. Correspondence concerning this article should be addressed to Elana Joram: Department of Educational Psychology and Foundations, University of Northern Iowa, Cedar Falls, IA 50614-0607 or by Internet to Joram@uni.edu.

provides information about principles of measurement–for example that units must be of equal size and contiguous (Hiebert, 1981; 1984)–in the absence of physical reminders the estimator must have knowledge of these principles and apply them. Instruction on measurement estimation then, can serve as a convenient conduit for teaching the principles of physical measurement: a central topic in the elementary mathematics curriculum and a building block for other mathematical concepts, such as ratio, fractions, and geometry (Coburn & Shulte, 1986; Davydov & Tsvetkovich, 1991; National Council of Teachers of Mathematics, 1989). A corollary is that research on measurement estimation has the potential to reveal fundamental understandings and misunderstandings about physical measurement.

In addition to serving as a means of teaching and assessing understanding of physical measurement, measurement estimation is also an important competency in its own right and is often the only method for solving problems that is available or needed (Levin, 1981; O'Daffer, 1979). Usiskin (1986) comments that although estimation is often considered the "weak sister" to mathematical computation, in real life it is more frequently the "stronger sister" or the "only child." Because estimation does not require tools, estimators can perform their tasks quickly, and for tradespeople such as carpet layers (Masingila, 1993) and seamstresses (Hancock, 1994) this may lead directly to monetary gains.

Because measurement estimation forms a foundation for understanding physical measurement and is also an essential numeracy skill (MacNeal, 1994; Paulos, 1988), it has been identified as a critical area for development in mathematics education (National Council of Teachers of Mathematics, 1989). Interest in all types of estimation has grown recently among mathematics educators in hopes of combating the well-documented failures of American students to approach mathematics as a meaningful domain (Beaton, et al., 1996; National Council of Teachers of Mathematics, 1989; Shaughnessy, Nelson, & Norris, 1997; Sowder, 1992). When estimating, students are unlikely to mindlessly apply memorized but ill-understood procedures, and so successful estimation makes likely the application of number sense (Hiebert, 1981).

Despite its recent central position in the mathematical playing field, however, research on measurement estimation is sparse (Sowder, 1992). According to Sowder (1992), in all areas of estimation "there simply is not a rich research base" (p. 372) because, although investigators agree on the importance of this topic, they do not agree on which research issues ought to be pursued nor how they should be examined. We add to this list, as perhaps the leading cause of this state of affairs, a lack of theoretical models of the measurement estimation process that would serve to guide researchers. For all of these reasons, measurement estimation is a topic that continues to be neglected by researchers.

Those studies which have been conducted on measurement estimation performance (see Sowder, 1992) portray a dismal picture: Children and adults alike are poor at estimating, with typically fewer than half of participants in any study achieving a criterion level of accuracy. These studies provide little insight into how people make their judgments, and why measurement estimation seems to be so problematic even for adults. A similarly incomplete picture emerges from an examination of studies on instruction in measurement estimation: Until recently, instructional interventions were based on behaviorist assumptions about the importance of practice with feedback in learning. Although these methods have been

partially successful (e.g., Ferris, 1972; Jones & Rowsey, 1990), we are left with little understanding of why the methods work under certain conditions but fail to work under others. In order to make some headway in addressing the issues described above, we review existing literature on performance and instruction in measurement estimation through the lens of a framework for the measurement estimation process.

We begin by describing the conceptual prerequisites of measurement estimation, focusing on competency in physical measurement. We then ask what is needed for measurement estimation beyond this knowledge. Educational studies of performance in measurement estimation are briefly reviewed, and several models of the measurement estimation process are discussed. We then propose an alternative framework and discuss how estimation strategies can be interpreted within this framework. We briefly review studies of educational interventions for measurement estimation, and analyze them in light of the proposed framework. Finally, based on predictions generated from our framework and on conclusions drawn from the review of instructional research, we make recommendations for future directions in research and instruction.

In this review, we focus on the estimation of linear measurement, such as length and distance, because most research on measurement estimation has been carried out on these attributes. The cognitive processes involved in estimating will likely vary depending on the attribute estimated, and our framework is designed to account for those processes involved in estimating linear measurements. We include studies in which participants estimated linear measurements as all or part of the assessment. Finally, we restrict our review to studies in which participants were asked to use numbers to make their judgments because we focus on people's understanding of the relations between perceived magnitudes and their numerical representations.

Conceptual Prerequisites of Measurement Estimation

The conceptual prerequisites for measurement estimation fall into two categories: logical reasoning processes described by Piaget (Piaget, Inhelder, & Szeminska, 1960), and knowledge of specific measurement concepts (e.g., Hiebert, 1981, 1984). Two logical reasoning processes have been considered to be closely related to linear measurement: length conservation and transitivity (Hiebert, 1984). Both are thought to develop in children by 6 to 7 years of age (Boulton-Lewis, 1987).

Hiebert (1984) notes that a popular belief is that children cannot learn measurement concepts until they have fully developed conservation and transitivity. Hiebert (1981) demonstrated, however, that first grade children who had not yet developed these processes could be successfully taught some physical measurement concepts. He suggests that the absence of conservation and transitive reasoning does not appear to limit children's learning of most measurement concepts. We can, therefore, conclude that by first or second grade most students will be capable of learning physical measurement concepts.

Results of educational studies of children's performance on physical measurement tasks indicates that despite their cognitive readiness to learn measurement concepts, children have considerable difficulty acquiring these concepts. By fourth grade, students can perform "low level" tasks well, such as recognizing units and measuring with a ruler in customary units. However, when presented with more challenging tasks, many sixth and seventh graders continue to give incorrect answers (Carpenter, Kepner, Corbitt, Lindquist, & Reys, 1980; Hiebert, 1981; Kouba et al., 1988). For example, Hiebert (1981) found that only 50% of fourth-grade students and 62% of sixth-grade students could find the number of paper clips needed to "cover" the length of a drawn line segment. Carpenter et al. (1980) and Hiebert (1981) interpret such errors as reflecting gaps in students' understanding of basic concepts of linear measurement such as "measuring is covering with units." In summary, the discussion above suggests that although children have the capacity to understand linear measurement concepts by early elementary school, basic concepts of physical measurement continue to be challenging for many students through middle school and beyond.

We can hypothesize that if understanding the concepts of physical measurement is a prerequisite for deploying these principles mentally in measurement estimation, misconceptions about physical measurement will be evident when estimating. For example, if a student has not learned the unit-covering principle, they may mentally iterate units to estimate length, yet fail to completely cover the length of the extent they are estimating. A related point is that because students' knowledge of the concepts of physical measurement increases as they advance in grade level, improvements in estimation accuracy with age may be related to this growing base of knowledge of physical measurement. These hypotheses remain speculative, however, and require research investigating the relationship between knowledge of physical measurement and measurement estimation skills. We have suggested above that knowledge of specific concepts of physical measurement is one prerequisite for measurement estimation. We now turn to the question of what additional cognitive processes might be involved in measurement estimation beyond those required for physical measurement.

Is it simply that measurement estimation is a process that is isomorphic to physical measurement, and that if one can measure one can estimate? This would suggest a straightforward answer to the question of why measurement estimation is difficult: Estimators simply do not understand how to measure. However, we suggest that, although the conceptual bases for physical measurement and measurement estimation are identical, the processes by which an estimator arrives at an estimate or constructs a mental representation of a particular measurement are not. One who is measuring with an instrument needs to know certain procedures: for example, to line up a ruler so that it is flush with the extent to be estimated. An estimator, on the other hand, may make use of a variety of strategies to cope with the conceptual demands of estimation, and these too require expertise. Lending support to this point, Corle (1960) found that fifth and sixth grade students who first estimated and then measured the items they had estimated made estimating errors about twice as often as they made measurement errors. It is the layer of knowledge over and above an understanding of physical measurement that is the focus of our review. We begin by describing three generations of studies of measurement estimation performance that have been carried out in education, to establish what is currently known in this field about how people estimate linear measurements.

Educational Studies of Measurement Estimation Performance

Educational studies of measurement estimation performance have been reviewed elsewhere (Sowder, 1992), and are discussed only briefly here (see

Appendix A). In the first generation of studies (Cassell, 1941; Crawford & Zylstra, 1952; Wilson, 1936; Wilson & Cassell, 1953), researchers were interested in simply documenting how well children and adults could estimate. Individuals were asked to estimate measurements of remembered, familiar objects such as the height of a dining room table, leaving open the possibility that the estimated objects varied in magnitude from individual to individual[1]. Responses were classified as "reasonable" or "incorrect," and scoring was relatively strict, with a narrow band of estimates permitted for correct responses. Children and adults performed poorly in these early studies: Although there were improvements in accuracy with increasing age, reasonable estimates were given by only about 30 to 40% of high school seniors, and fewer than 50% of adults, on average.

In the next generation of studies (Corle, 1960, 1963; Swan & Jones, 1971, 1980), researchers focused on differences in people's ability to estimate various attributes such as height and weight. An advance over earlier studies was that the objects to be estimated were physically present as people estimated instead of being remembered, ensuring that different individuals estimated the same objects. Percent error[2] was introduced as a measure of estimation accuracy, with percent error scores sometimes classified into categories such as "reasonable," or "incorrect." Although increases in accuracy with age were again observed, the average accuracy level of children and adults in these studies was low and performance was found to vary widely across attributes. Estimates of linear measurements were generally found to be more accurate than for weight, liquid capacity, volume, but poorer than for temperature.

The most recent group of studies of measurement estimation (Forrester, Latham, & Shire, 1990; Forrester & Shire, 1994; Siegel, Goldsmith, & Madson, 1982), conducted in the '80s and '90s, shows the influence of a prevailing interest in cognitive processes in the field of education. Rather than focusing only on participants' accuracy levels, researchers began to look beneath the surface to try to figure out how people make their judgments. This they did by manipulating a variety of variables hypothesized to influence measurement estimation, such as the need to mentally decompose an object prior to estimating, and by probing participants to find out what strategies they had used. Objects were again physically present during estimation, and two methodological features, borrowed from the field of psychophysics, were introduced: the use of power functions[3] and nonstandard units. The studies revealed that measurement estimation is a highly volatile process, and easily influenced by features of to-be-estimated objects. Results of these studies were used to advance process models of measurement estimation.

Forrester et al.'s (1990) model identifies the background abilities, "specific problem-space" skills, "social-discursive context" skills, approximation skills, and procedures that all bear on measurement estimation. In the flowchart model proposed by Siegel et al. (1982), the estimator first checks to see if a benchmark can be applied, and if unavailable, proceeds to subdivide the to-be-estimated object. These two models begin to give us a sense of the many different elements that may be involved in generating estimates, and take us beyond the simple reports of accuracy levels that characterizes previous educational research in this area. The models of the estimating process that have been proposed in education, however, have several limitations. First, as Sowder (1992) notes, they fail to distinguish between the estimation of numerosity (discrete quantities) and measurement estimation (continuous quantities). Second, the models are unwieldy, and include many possible steps or elements, yet descriptions are at a global level and provide an unsatisfying account of what cognitive processes estimators engage in while estimating. Finally, the models have not been linked to instruction; consequently, most instructional interventions designed to teach measurement estimation reflect the prevailing zeitgeist of psychology or education rather than specific ideas about what needs to be promoted in measurement estimation.

In the next section, we present a framework that distinguishes between numerosity and measurement estimation, but shows how they are conceptually related. The framework is more parsimonious than models previously proposed, yet at a finer-grain level of description. It is intended to make some inroads into providing a basis for understanding measurement estimation, for evaluating methods of instruction, and for making recommendations for further instructional work in this area.

A Framework of the Measurement Estimation Process

Of the models of the measurement estimation process that have been suggested in both education and cognitive psychology (Forrester et al., 1990; Siegel et al., 1982; Thorndyke, 1981), only one is at a sufficiently detailed level to make it relevant to the present discussion. Thorndyke (1981) proposed the *analog timing model* to explain how people estimate distances on maps. In his model of distance estimation, individuals scan a distance on a map and keep track of the time taken to do so. They then scan a single unit, and estimate total distance by dividing total scanning time by the time taken to scan one unit–distance estimates are thus inferred from time estimates. The final step in the process, when working with maps, is to multiply the number of units by the number of miles that each unit represents.

Of particular interest to Thorndyke (1981) was explaining the well-established effects of "clutter" in cognitive psychology–that distances filled with many objects, points, or landmarks, are judged as greater than unfilled distances (e.g., Kosslyn, Pick & Fariello, 1974; Luria, Kinney & Weissman, 1967; Pressey, 1974). This he did by positing that for each point encountered on a map in his task, an individual stops the scan, retrieves the name of the city, and compares it to the name of the final destination–the scan continues if the names do not match. Total scanning time, and therefore distance estimated, would increase for each point encountered during the scan, and Thorndyke's (1981) experiments confirmed these predictions.

Thorndyke's (1981) model of distance estimation appears to apply well to the specific task given to participants in his study. It does not, however, apply as well to other estimating tasks, for example, those in which participants do not have to stop to retrieve the names of cities while estimating. This model is also inconsistent with self-reports of estimators in which the majority describe mentally segmenting lengths and distances into units as they estimate (Forrester et al., 1990; Friebel, 1967; Hartley, 1977; Hildreth, 1980, 1983; Immers, 1983), and with other measurement estimation strategies that have been observed (Forrester & Shire, 1994; Friebel, 1967; Hildreth, 1980, 1983; Moskol, 1980; Siegel et al., 1982). Below, we propose a framework for the estimation process that encompasses a variety of cognitive processes, identified during measurement estimation, and accounts for the results of a wider range of studies than previous models.

Bidirectional Mapping: From Measurement to Quantity and Back

Gallistel and Gelman (1992) proposed a model of verbal and nonverbal counting in humans, which we now apply to measurement estimation. They build on Meck and Church's (1983) accumulator model of the process by which nonverbal animals obtain magnitude representatives of numerosity. A mental magnitude is a signal in the brain that is formally equivalent to a quantity, such as the amount of water in a graduated beaker. In the Meck and Church model, animals count nonverbally by, in effect, pouring a cup of water into a graduated beaker for each item in the series or array that is being counted. The resulting accumulation is the magnitude that represents the numerosity of the counted set, just as the height of a bar in a histogram represents a count. In the Meck and Church model, although the process that generates a cardinal value is discrete, the representation is a continuous quantity. Given this, the representation has the characteristics of other continuous quantities such as interval durations, line lengths, and so on.

Like other mental magnitudes, the magnitudes that represent numerosity have scalar noise. That is, the values obtained from reading a magnitude in memory vary from one reading to the next and the width of the distribution of such readings varies in proportion to the magnitude. In Figure 1, the variability in the magnitudes to which the digit at the bottom of the column maps is shown by the fading from black to grey at the top of a column. Figure 1 shows that the interval over which a column fades increases in proportion to the mean magnitude. Consequently, the discrimination of such magnitudes obeys Weber's law, that is, the discriminability of two magnitudes is determined by their ratio. In other words, Gallistel and Gelman (1992) predict that the discriminability of two numbers decreases as their mean numerical value increases because the variability in the mapping is scalar.

Gallistel and Gelman (1992) suggest that the nonverbal counting process provides a model and interpretation for verbal counting, which facilitates the learning of verbal counting. Moreover, they suggest that nonverbal counting goes on in parallel with verbal counting, so that, when the learner uses verbal counting to obtain the verbal symbol for numerosity, the nonverbal counting process simultaneously delivers the nonverbal mental magnitude that represents that numerosity. The parallel operation of nonverbal and verbal counting enables the child to learn a *bidirectional mapping* between the verbal symbols for numerosity and the corresponding nonverbal symbols (the magnitudes). As the bidirectional mapping develops, so does a meaningful understanding of the symbols we call the natural numbers. This mapping, which functions automatically in adults, retrieves from memory the magnitude that corresponds to a given number word that represents the generated numerosity, and it automatically retrieves from memory the number words that might correspond to a given magnitude. Because the magnitudes have scalar noise, this latter process does not always retrieve the same count word for a given mental magnitude; rather it retrieves a word "in the right ballpark."

This bidirectional mapping means that in the process of learning to count verbally, humans construct a mental number line and assign the verbal symbols positions along that line. Whalen, Gallistel, and Gelman (1999) have recently provided experimental evidence for a high-speed nonverbal counting process in adults. They demonstrated that adults could, with reasonable accuracy, assess the

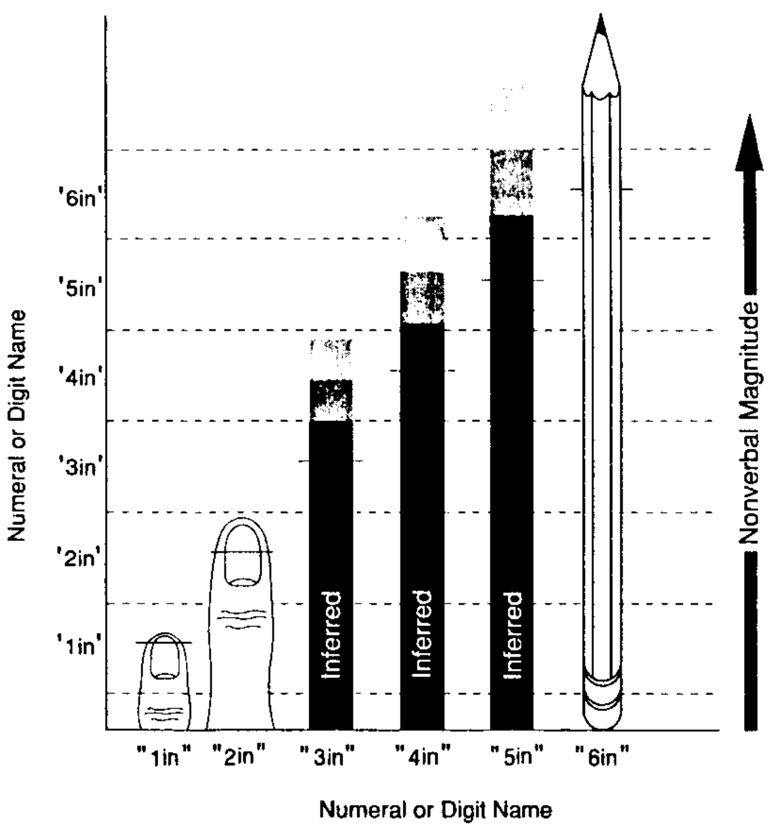

FIGURE 1. *Mental measurement line for inch unit, showing how reference points can embody nonverbal magnitudes. Nonverbal magnitudes are shown by the objects and grey columns. The variability in nonverbal magnitudes is represented by the fading at the tops of the grey columns. Written or spoken measurements (numerals or digits) are indicated by the numbers in quotes along the bottom of the graph. The mapping from the nonverbal magnitudes to digits, shown on the left, is effected by subdividing the mental number line into segments labeled by measurement (numerals or digits). Adapted from Gallistel and Gelman (1992).*

numerosity of quantities up to 25 when verbal counting was precluded due to time limits. Their studies provide empirical support for the notion that adults have a nonverbal counting mechanism and that they represent numbers as nonverbal magnitudes. Their evidence also implies a bidirectional mapping between verbal (or written) number symbols and mental magnitudes with scalar variability.

The numerically competent individual, according to Gallistel and Gelman (1992), will have learned to map from digits to nonverbal magnitudes: for example, upon seeing a set of 6 objects he or she will count these objects verbally and use that number to find its corresponding nonverbal magnitude on a mental number line. In addition, he or she will be able to perform the inverse mapping from nonverbal magnitudes to digits: This entails subdividing the range of nonverbal magnitudes and assigning to these intervals their appropriate digits. In other words, the individual will have constructed a mental number line which permits access to the corresponding natural number. This process, which is often called subitizing, provides an alternate, and quicker, way to obtain the word that represents the numerosity of a small set (Gallistel & Gelman, 1991).

Applying Gallistel and Gelman's (1992) ideas to measurement estimation, we hypothesize that humans can obtain a verbal or written numerical representation of a magnitude in three ways, all of which are available to adults: (a) the direct verbal way, when one mentally divides the magnitude under consideration into discrete intervals and verbally counts the intervals; (b) the indirect way that involves dividing the magnitude into discrete intervals, counting the intervals nonverbally to obtain a mental magnitude that represents numerosity, and then using the bidirectional mapping to obtain a verbal representative of the corresponding numerosity; and (c) the indirect way that is used when the magnitude (e.g., a length) directly generates a nonverbal magnitude (a perceived magnitude) for which a mental scaling factor has been learned that relates that magnitude to the magnitudes that represent numerosity. The latter way is only available to someone with repeated estimating experiences of the first two kinds, because it is only those kinds of experiences that can establish the scaling factor that relates perceived magnitudes to numerical magnitudes.

As with research findings on numerosity judgments, we would predict that reaction times to estimate to-be-estimated items (referred to hereafter as TBEs) would increase linearly with the number of units to be estimated, if estimators are using the first hypothesized process above. This prediction is confirmed by Hartley's (1981) research, where he found that average correlations between reaction time and number of units estimated ranged from .70 to .86 across four experiments. The third process may occur when estimators use strategies in which they recall a familiar object whose measurement is known and compare it to the TBE. For example, when estimating the width of a room that is 12 feet, an estimator may recall the length of a crocodile they know to be 12 feet long (Hildreth, 1980). In this case, the crocodile stands for the nonverbal magnitude and provides an immediate entry point to it on the mental number line. Rather than segmenting the room into 12 foot-long units, the estimator has reduced the number of units to be counted to 1, represented by the crocodile.

The processes described above may also be used in a second kind of measurement estimation described by Bright (1976), which requires an individual to select an object that matches a given measurement on some attribute (e.g., finding an object that is about 6 inches in length). In such tasks, or those that require an individual to produce a physical magnitude (e.g., draw a 6-inch long line), the estimator may generate magnitudes on the mental number line in the moment, based on their knowledge of the size of an individual unit and the number of units they are to mentally represent. For example, upon hearing "6 inches" the estimator mentally concatenates 6 of what she or he knows to be an inch-long magnitude until the TBE is fully represented. Alternatively, individuals who have learned the mapping between the measurement and its corresponding nonverbal magnitude can use the verbal or written representation (6 inches) to find the corresponding nonverbal magnitude on a number line. The nonverbal magnitude would then provide the basis for either selecting an object of the same magnitude or drawing a physical representation of the nonverbal magnitude.

If we assume that Weber's law also applies to measurement estimation, we would predict that the greater the number of units needed to estimate, the greater the errors in estimation. Supporting this prediction, Campbell, Fein, and Schwartz (1991) found that the mean accuracy of students' estimates dropped significantly as distances to be estimated increased and the units remained the same. Other predictions based on the application of this law to measurement estimation have yet to be investigated. For example, the "direct entry route" to the number line, like subitizing, may be used only for estimating a small number of units. The rationale behind this prediction is that as the number of units increases, the likelihood of error also increases and this can be offset by using the more accurate subdivide and count strategy.

How units are counted during estimation.

Whalen et al. (1999) offer the following "as-if" account of the proposed accumulator mechanism for counting. When an animal counts, it is as if it mentally pours one cup per counted item into a graduated beaker. As each cup is added to the column, the quantity increases by one. When a cup has been poured for each item to be counted, the final quantity represents the nonverbal cardinal value. This model can be applied to measurement estimation with a few modifications. The main difference between counting and measurement estimation is, obviously, that counting applies to discrete quantities whereas measurement estimation applies to continuous quantities. We suggest that in measurement estimation the estimator renders the continuous TBE or its remembered image discrete and then engages in the same processes as featured in the counting model.

Gallistel and Gelman (1992) note that when counting the numerosity of a set, the representatives of numerosity are continuous magnitudes. When counting, these continuous nonverbal magnitudes are rendered discrete by means of the single burst of impulses to the accumulator for each item in the set. The discrete items themselves supply the perceptual cues as to when each burst should occur. We suggest that what makes measurement estimation significantly more challenging than counting is that the estimator must mentally supply the boundaries that indicate where, and therefore when, each burst should occur. Using the metaphor above, a further complicating factor is that the division points change depending on the size of the cup (unit) that is used.

When estimating a continuous quantity, we might hypothesize that the estimator sets the amount of time to gate each burst according to the size of the unit that

is used. Thus, when estimating in inches, a relatively short duration will be used to gate each burst; when estimating in yards, a longer duration will be used. Alternatively, we might suppose that the estimator will count units as when counting discrete objects, and that the actual length of the unit is irrelevant to the time required to do so. The latter hypothesis is supported by Hartley's experiments (1981) in which adults estimated line lengths using a 1-cm standard (unit) or a 2-cm standard. Lines that were estimated were the same multiples of the different units, for example, participants estimated a 6-cm line in 1-cm units, and a 12-cm line in 2-cm units. Hartley found that reaction time was well accounted for by the number of "laying offs" to be done rather than by absolute line length–in the example above, reaction times were the same for estimating both lines. These results surprisingly suggest that when both forming and manipulating mental images, time to estimate is accounted for by the number of laying offs to be done, rather than by actual distance estimated or by the magnitude of the unit.

Kosslyn, Ball, and Reiser (1978) found that the times required to scan mental images increased with distance, and that images which were subjectively larger required more time to scan than those which were smaller, seemingly contradicting Hartley's (1981) claims. Similar results were obtained by Kosslyn, Reiser, Farah, and Fliegel (1983), who found that more time was required by adults to scan between two imaged objects if they were pictured at greater distances. However, a systematic relation between distance and reaction time was only found under conditions where participants were instructed to scan their images, for example to image "a little black speck zipping in the shortest straight line from the first object to the second" (Kosslyn et al., 1978, p. 52). When participants were not given instructions to scan images but merely to make a decision about whether a named object was on an imaged map (Kosslyn et al., 1978), or to determine the distance of objects from a starting point (Lea,1975), response times have been found to be unrelated to distance scanned.

The effects of "clutter" on estimation.

Thorndyke (1981) argued that longer reaction times to estimate spaces that are cluttered (filled with objects) result from the estimator retrieving and comparing each item encountered with an end state in order to determine whether they had visually reached their final destination. Thorndyke claimed that such findings support his model of the estimation processes in which distances are first scanned and then total time taken to scan is divided by the time taken to scan one unit. However, this explanation does not account for longer reaction times to estimate cluttered distances in studies where no such retrieval and comparison would be predicted. For example, Luria et al. (1967) found that participants judged the distances of objects to be greater when the space between the participant and the object was cluttered than when the space was uncluttered. Kosslyn et al. (1974) also found that adults perceived distances between toys that were separated by an opaque barrier as longer than distances separated by a transparent barrier or by no barrier at all.

A study outside the domain of linear measurement may shed some light on why clutter results in estimates of greater magnitude. Participants in a study by Poynter (1983) estimated durations in which three "high priority events" were either clustered at the beginning of an interval or distributed throughout an interval.

TABLE 1
Strategies for measurement estimation

Strategy	Synonyms	Description	Examples
Mental Measurement			
1. Mental instrument application (Friebel, 1967)		Mentally lines up a measuring instrument with TBE and reads off measurement.	Mentally lines up ruler along TBE and reads last number to get estimate (Friebel, 1967).
2. Unit iteration (Hildreth, 1980; 1983; Immers, 1983)	Unit application (Friebel, 1967) Laying-off a standard (Hartley, 1977)	Mentally segments TBE into units, counts units, reports sum as the estimate.	
Transform the Unit			
1. Recall reference point	Benchmarks (Moskol, 1980) Unit application (Friebel, 1967)	Recalls an object equivalent to a standard unit, then applies another strategy such as unit iteration, with object substituting for standard unit.	To estimate height of a small building, recalls that an average doorway is about 7 ft. tall (Spitzer, 1961).
2. Prior knowledge (Hildreth, 1983)		Uses prior knowledge about part of TBE.	Counts ceiling tiles known to be 1-foot long to get estimate of room length (Hildreth, 1983).
3. Comparison (Hildreth, 1983)	One-to-one correspondence (Moskol, 1980)	Recalls reference point, whose measurement is known, that is equivalent in magnitude to TBE.	Recalls image of 12-ft. alligator, and uses it to estimate length of 12-ft. room (Hildreth, 1983).
Transform the Object			
Decomposition /recomposition	Chunking (for decomposition only) (Hildreth, 1983) Indirect method (Moskol, 1980)	Prior to estimating, subdivides TBE into smaller continuous quantities. Estimates are obtained for each part, using one of the processes above. Estimates for parts are then summed or multiplied to get final estimate.	Mentally divides a parking lot into spaces, estimates width of each space by comparing it to one's own height, then multiplies this estimate by number of spaces to get final estimate (Siegel et.al., 1982).
a. Use of subdivision cues (Hildreth, 1983)		Uses physical aspects of TBE to mentally subdivide.	Uses painted lines in parking lot as cues for subdividing (Hildreth, 1983).
b. Rearrangement (Hildreth, 1983)		When subdivision is difficult (irregular decomposition) estimators mentally rearrange part of TBE prior to subdividing.	To estimate length of L-shaped line segment, mentally moves one line to make a straight line.
Other			
Sequencing (Hildreth, 1983)		By making estimates that are a little more and a little less than the true measurement, the estimator gets progressively closer to the true measurement.	To estimate the length of a 70-cm. line, the estimator first estimates it as between a half meter and one meter. The estimator uses another strategy, such as comparison, to complete the estimate (Hildreth, 1983).

Duration estimates were greater when the events were distributed throughout the interval, and the author concluded that the greater the number of "segmenters of experience," the greater the duration estimates. We can conjecture that more clutter provides prompts to gate extra bursts to the accumulator, resulting in estimates of greater length or distance than when less clutter is present. However, studies need to be carried out in the domain of linear measurement to confirm this hypothesis.

Segmenting a continuous quantity into units.

Counting does not appear to present conceptual difficulties for individuals, and researchers (e.g., Gelman, 1991) have suggested that infants are innately prepared to learn to count. If individuals rely on similar processes when estimating continuous quantities as when counting discrete objects, then what would explain the difficulties people have on estimation tasks? As described above, when segmenting a TBE and counting units, the estimator must mentally transform the continuous quantity into discrete quantities that can be counted–in other words, they have to know where to segment the quantity into units. This is no trivial task in that it requires: (a) recalling an image of the unit if not present or constructing an image if the unit is present; (b) repeatedly comparing an image of the unit to the TBE; (c) keeping track, when laying off the unit, of where the last unit ended and where the next one must begin, and finally, (d) maintaining a running tally of the number of units while simultaneously performing the previous tasks. When we add knowledge of measurement principles to this list, such as knowing units must be contiguous and equal in size, measurement estimation becomes a formidable task, indeed. Estimators invent a variety of strategies to cope with these demands (e.g., Forrester & Shire, 1994; Friebel, 1967; Hildreth, 1983; Moskol, 1980; Siegel et al., 1982) that go beyond iterating a unit. We discuss some of these strategies, as well as *unit iteration*, in the next section.

Strategies for Measurement Estimation

Our discussion of measurement estimation strategies is organized into three major categories: mental measurement, transforming the standard unit with which estimations are made, and mentally transforming the TBE prior to estimating (see Table 1). We focus, in this discussion, on unit iteration and the use of reference points.

Mental Measurement

Mental measurement includes imagining a measuring instrument such as a ruler lined up with the TBE, as well as the strategy of unit iteration. When using the unit iteration strategy, according to the model above, the estimator segments the TBE into units, verbally counts the number of units, and then finds the place on the mental number line that corresponds to the total number of units counted. Iterating a standard unit has been found to be the most commonly used strategy for measurement estimation in all studies that have asked individuals to report their strategies (Forrester, 1990; Friebel, 1967; Hartley, 1977; Hildreth, 1983; Immers, 1983). The percentage of participants who use the unit iteration strategy ranges from approximately 30% (Forrester, 1990; Friebel, 1967) to 97% (Hartley, 1977), across different studies.

In an exploratory study, Hartley (1977) found that 97% of adults reported that they had formed a mental image of a nonstandard unit, which was given by the experimenter and was present while they estimated, to estimate the lengths of lines. To empirically support his participants' self-report data, Hartley (1981) subsequently showed that the time adults took to estimate the length of a line increased linearly with the length of the lines. Hartley (1981) interpreted these findings as evidence that adults were "laying off a standard," in other words, mentally iterating a unit, in order to make their estimates. Immers (1983) also found supporting evidence for the use of this strategy: He noted that of the 81% of elementary students in his study who reported using a unit iteration strategy, about one-third marked off units physically, and the other two-thirds were observed to mark off units with their eyes.

Strategies That Transform the Unit

The unit iteration strategy requires estimators to recall the unit if it is not present during estimation, and to segment the TBE into units; both pose cognitive challenges for estimators. To cope with these difficulties, estimators may transform the unit, in the manner described below.

Reference points.

Instead of comparing the TBE to a standard unit, the estimator may make use of a reference point which they then iterate in the same manner as the standard unit. For example, an image of a person who is known to be approximately 6 feet in height could be mentally iterated across a room to get an estimate of the room's width. One rationale for using reference points is that they increase the likelihood that the mental representation of unit size will be accurate because the estimator can presumably more easily represent the reference point than the corresponding standard unit. A second reason for using reference points is that they may be more meaningful to the estimator than standard units. For example, Carter (1986) suggests that when a child asks "How far is it to the public library?" a response of "About as far as the park" will be more meaningful to the child than "three-quarters of a kilometer" because the distance to the park is known to the child. Finally, reference points that are greater in magnitude than individual standard units reduce the number of iterations that must be performed. For example, Spitzer (1961) notes that a person who has a reference measure for 50 miles can quickly understand the statement "The village is 90 miles from the Canadian border" (p. 235). In the strategy known as "comparison" a single reference measure which is equivalent to the magnitude of the TBE is used, thus obviating the need for any iterations at all.

Whereas there has been much discussion in the measurement estimation and number sense literature about the importance of building a set of meaningful reference points for measurement estimation (Bright, 1976; Carter, 1986; Crawford & Zylstra, 1952; Duncan & Litwiller, 1986; King & Whitman, 1973; Markovits, Hershkowitz & Bruckheimer, 1989; Neufeld, 1989; Siegel et al., 1982; Spitzer, 1961; Swan & Jones, 1980; Thompson & Rathmell, 1989; Van de Walle & Thompson, 1985) there have been few investigations of the extent to which estimators, skilled or otherwise, make use of such reference points. Nonetheless, current textbooks (e.g., Scott Foresman/Addison Wesley, 1998) encourage elementary level students to work with reference points for standard units of measurement.

What estimators gain by using reference points.

Clearly, using reference points in some of the cases described above reduces many of the cognitive demands of estimating above and beyond easing the recall of a standard unit: because the number of unit iterations is reduced, the estimator less frequently needs to keep track of where one unit ends and another begins, and there are fewer units to tally.

In addition to the practical significance of using reference points to deal with the cognitive demands of estimating, constructing a set of reference points may help the estimator develop an internal scale for an attribute, with different reference points positioned appropriately along the scale. Such a scale, according to the model of the estimation process described above, would allow estimators direct access to nonverbal magnitudes. In fact, we might imagine that an individual very facile in the estimation of a particular attribute and unit would have different reference points positioned along their mental number line that embody continuous quantities. On a mental number line for inches, one might have a thumb from knuckle to finger tip in the "1" location, an entire thumb in the "2" location, and a pencil in the "6" location (see Figure 1). The other magnitudes may be constructed "on-the-fly" during the estimation process based on inferences generated from the known reference points. In other words, a *mental measurement line* need not have all reference points filled in because those missing can be generated based on the estimator's knowledge of how a scale is constructed. We can predict that when individuals learn a set of reference points for specific measures (e.g., 1 foot, 5 feet, 10 feet), they should be able to more rapidly identify intermediate measurements for which a specific reference point was not learned (e.g., 7 feet), than measurements that extend beyond the bounds of a scale of the learned measures, or which are on another scale.

An individual who lacks facility in measurement estimation may have only a mental number line and a representation of the magnitude of a single unit (e.g., 1 inch)–when estimating, the individual must mentally iterate that unit and thereby construct the nonverbal magnitudes that are otherwise not available. We suggest this describes the situation of most estimators–hence, the finding that the most common strategy is unit iteration. Some children do not correctly represent even a single unit, ensuring that estimates generated through the process described above will be inaccurate. Number sense in measurement (Markovits et al., 1989; Sowder, 1992), can be understood as having a mental measurement line on which nonverbal magnitudes are represented for different units. Reference points provide a means of constructing such a line in a way that is meaningful and memorable for the estimator.

Strategies that Transform the To-Be-Estimated Item

Decomposition/recomposition.

Rather than transforming a standard unit to facilitate subdivision of the TBE into units, an estimator may mentally decompose the TBE into smaller continuous quantities prior to estimating and then apply one of the processes described above. *Subdivision cues*, present in the TBE, may suggest ways that the TBE may be decomposed, for example, lines on ceiling tiles which may be used to estimate the width of a room. Estimates obtained for smaller parts of the TBE are then added or multiplied in a process known as *recomposition*. When TBEs cannot be easily subdivided into approximately equal parts (*irregular decomposition*), estimators may mentally rearrange the TBE prior to estimating (Hildreth, 1983). Such strategies suggest that estimators may nonverbally measure TBEs prior to estimating, or at least perform a quantitative assessment of the TBE.

Discussion of Research on Measurement Estimation Strategies

The Effects of Task Demands on Estimation Strategies

Mental measurement is the most common strategy to emerge in studies that have investigated measurement estimation strategies (Friebel, 1967; Hartley, 1977; Hildreth, 1983; Immers, 1983), but these findings must be considered in light of the special demands of the tasks used by the researchers: in all cases, the TBEs were lines less than 11 inches long. An unexplored issue is whether people's strategies for measurement estimation vary depending on the magnitude of the TBE. For example, we can conjecture that when estimating a relatively long real- world object, an individual would be likely to use a reference point in order to reduce the number of unit iterations that would need to be performed. Research is needed to document the manner in which estimators' strategies interact with task demands such as size of the TBE and whether units and TBEs are present during estimation.

The Relationship of Age to Estimation Strategies

A major gap in the literature on measurement estimation strategies is the lack of research on the relationship of the age of the estimator to strategies deployed. Virtually all strategies discussed above rely on mental imagery to both generate and manipulate images of units and/or TBEs. Research by Kosslyn, Margolis, Barrett, Goldknopf, and Daly (1990) suggests that it is not until around third grade that students can successfully generate and mentally manipulate images. Increased facility with mental images even beyond third grade may play an important role in explaining age-related increases in estimation accuracy. For example, Forrester, Latham, and Shire (1990) found that the use of the unit iteration strategy increased with age, as did estimation accuracy. Further, the mediating effects of imagery on estimation accuracy may be more pronounced in tasks that require estimators to imagine the unit (or reference point), the TBE, or both. If development brings increased facility with mental imagery and perhaps the assembly of a repertoire of familiar reference points, we would predict that the effects of task demands will diminish with increases in the age of estimators.

Instructional Studies on Measurement Estimation

Two main types of instruction have been used to teach measurement estimation: one is the *guess and check* method (Bright, 1976) or *practice with feedback* (Ferris, 1972), and the other is *strategies training* (Hildreth, 1983). Recently, a third kind of instruction has been attempted, focusing on number sense in measurement estimation (Markovits, 1987), or what Bell (1974) calls *measure sense*. In this section, we briefly review studies falling into the first two main categories (see Appendix B), highlight the pattern of results, and discuss why interventions may

be effective in light of the framework of the estimation process described above.

The Guess and Check Method

In guess and check, the estimator first states or writes an estimate and then measures the TBE or is told the correct measurement (Bright, 1976). Most of the published research on instruction in measurement estimation, spanning a 70-year period, has employed the guess and check method (see Appendix B). These studies demonstrate that although guess and check is a powerful method for quickly improving accuracy, skills developed are fragile and limited to contexts identical or similar to those used in training. For example, Ferris (1972) found that SCUBA divers trained to estimate using the guess and check method in air transferred their skills to clear but not turbid water.

To determine what mechanism in the guess and check method is responsible for increases in estimation accuracy, Thorndike (1927) gave estimators feedback ("right" or "wrong") on the accuracy of lines they drew to a given length. He concluded that accuracy improves as a result of guess and check because feedback from the experimenter reinforces or weakens the estimator's responses. Trowbridge and Cason (1932) challenged Thorndike's hypothesis, and showed that the critical feature responsible for the efficacy of the guess and check method is the precision of feedback: In their study, a group that was told exactly how many eighths of an inch their drawing deviated from the desired line length achieved twice as many correct responses than a group that was simply told "right" or "wrong."

Studies conducted later suggest that guess and check improves estimation accuracy without the mediating impact of improved strategies for estimating. From an examination of the mean scores in Hildreth's (1980) study, we can see that Hildreth's guess and check group improved their estimation accuracy even though they did not increase the frequency of their use of strategies. Students in Immers' (1983) study also became more accurate in estimating after practicing with the guess and check method but did not alter their use of strategies from pre- to posttest. In summary, the results of studies discussed above suggest that the key element responsible for the efficacy of guess and check is informational feedback which allows estimators to hone their skills.

Strategies Instruction

There have been remarkably few published studies investigating strategies instruction for measurement estimation, particularly in light of the frequency with which strategies for estimating have been recommended in the mathematics education literature (e.g., Bright, 1976; National Council of Teachers of Mathematics, 1989; Spitzer, 1961; Thompson & Rathmell, 1989). The pattern of results evident from these few studies is far less clear than that from research on the guess and check method because most studies have conflated strategies instruction with the guess and check method, and have sometimes taught multiple strategies simultaneously (Hildreth, 1980, 1983). This makes it impossible to determine which aspect of the treatment was instructionally responsible for any observed gains in student performance and understanding. Several studies discussed below, however, suggest that the reference point strategy may enhance measurement estimation skills.

Attivo and Trueblood (1980) found that a group of preservice teachers who identified a personal reference point and then guessed and checked with the reference point were more accurate than students who constructed a reference point, removed it, and then guessed and checked. No differences on the posttest were detected between the personal reference point group and the pure guess and check group. The benefits of using a personal reference point disappeared, however, when participants were asked to estimate under conditions that differed from training. Without more information about how successful the preservice teachers were in Attivo and Trueblood's study at identifying personal reference points and how personal reference points differed from drawn or cut reference points, it is difficult to interpret the meaning of their results. However, findings from a study by Joram, Gabriele, Gelman, and Subrahmanyam (1996) also suggest that reference points are a promising method for teaching measurement estimation. Third-grade students in their study, who were taught to imagine reference points that were equivalent to an inch and a foot, drew more accurate representations of the corresponding standard units at posttest than a control group who had practiced the guess and check method. Estimation accuracy improved significantly more for the reference point group on certain items but did not transfer to real-world objects. The results of their study suggest that spatial manipulation of the image of a unit may play a large role in estimation accuracy, over and above the veracity of the image.

Discussion of Instructional Studies

Studies examining instruction on the use of strategies for estimating have produced a confusing pattern of results with respect to gains made on estimation accuracy; this pattern is certainly far less clear than that portrayed by use of the guess and check method. As discussed above, however, most studies that have investigated strategies instruction have suffered from several methodological flaws, and so it would be premature to conclude that strategies instruction is less effective than guess and check for teaching measurement estimation.

The gains found over and over again in studies using guess and check are puzzling–this form of instruction bears little resemblance to many of the reform-minded methods of instruction that are touted today. Instead, it seems to hearken back to behaviorist drill-and-practice methods of instruction. Why, then, does the guess and check method appear to be so effective? There are at least two possible answers to this question. One is that the guess and check method is not as effective as it first seems– it trains only a set of associations to a set of conditions but does not promote greater understanding of the principles underlying measurement estimation. The difficulty researchers have had in obtaining transfer in estimation skills under conditions that differ from those in training sessions, as well as the lack of strategies developed over the course of guess and check, suggests that this hypothesis is highly plausible. To assess its validity, studies are needed that examine the impact of the guess and check method on individuals' understanding of the principles and constraints that govern measurement estimation, as well as on their estimation accuracy.

The second possible answer to the question posed above is that the guess and check method actually helps students learn conceptual aspects of the estimation process. Developing numerical associations for a set of TBEs, as discussed above with respect to reference points, may enable the estimator to build their own *conceptual scale* (Gibson & Bergman, 1954; Gibson, Bergman, & Purdy, 1955)

for a given unit. Gibson and Bergman point out that such a scale might contain fractions and multiples of the basic unit. Gibson et al. felt that evidence for learning a conceptual scale through estimating with feedback was provided by posttesting that required participants to estimate the distance of unfamiliar targets in new locations. Learning the length of 3 inches, 5 inches, and 8 inches, for example, may allow the estimator to construct a representation akin to a mental number line for measurement, with points between whole numbers represented.

The mental measurement line, described above, would represent nonverbal magnitudes, either veridically, or perhaps through a function for each unit. For example, an individual may possess a basic number line much like that used for counting (see Griffin, Case, & Siegler, 1994; Resnick, 1983), which does not represent nonverbal magnitudes for different units. When asked to estimate in inches, for example, the estimator uses their image of a single inch to compute the needed nonverbal magnitude. Nonverbal magnitudes, according to this account, are constructed on-the-fly, rather than being precomputed and retrieved. Allowing for the idea of a generic mental number line that is converted to different conceptual scales for units through functions could explain why the size of units has not been found to be related to reaction times when estimating.

We suggest that part of the process of constructing a mental measurement line involves induction, where, based on the information given about points on a scale through feedback in the guess and check method, the estimator induces the size of an individual unit. Through deduction, the estimator then fills in points on the mental measurement line. For example, on the basis of feedback that certain objects are 2 inches, 5 inches, and 10 inches in length, the estimator can induce the magnitude of 1 inch and further hone the accuracy of this representation of a single unit through repeated feedback during guess and check. Transfer would be predicted to be most successful when the TBEs are on the scale that the estimator has conceptually constructed: that is, when estimating items that differ in magnitude from training stimuli but which are estimated on the same attribute and in the same units. This is exactly the pattern of performance that we observe in studies that have assessed participants trained in guess and check on their ability to transfer their estimation skills (Ferris, 1972; Immers, 1980; Morgan, 1986).

Where processes of induction may occur naturally through practicing guess and check, processes of deduction may need to be supported to both construct a conceptual scale and to extrapolate on this scale beyond points covered during instruction. This may be where the greatest limitation of the methods of both guess and check and reference points lies: Children learn to associate specific measurements with referents, but without being pushed to align these points in order along a continuum and to construct a mental measurement line, they may neglect to do so. This criticism is similar to those levelled at whole language methods for reading; whereas it is assumed that most children will induce phonetic patterns and syntactic rules from building a repertoire of associations between written symbols and their sounds, we now know that for many children this does not occur spontaneously (Smith, 1994). The argument can be made that the methods of guess and check and reference points target the semantics– the meaning level– of measurement estimation, but that what is also needed is instruction to target the syntactic aspects of measurement estimation–that is, the principles and constraints that underlie the construction and use of measurement scales for estimation[4].

When using the guess and check method, in some cases, estimators are permitted to check their own estimates using a measuring instrument. Estimators may thereby learn important principles of measurement and measurement estimation through interacting with measuring tools. That is, by using rulers estimators may become familiar with the size of a unit, and learn that units are iterated contiguously and that they completely cover the length of the TBE.

Thus, although the guess and check method of teaching estimation may at first seem like a conceptually barren drill-and-practice form of instruction, we suggest that it may be a very effective way to set up situations in which students can induce principles of measurement, and conceptual scales for units of measurement. Our analysis of the learning process that occurs during guess and check, however, suggests that this form of instruction needs to be supplemented because it does not invite the estimator to construct their own mental measurement line, particularly when measuring instruments are not used to check. Students will also need guidance in relating different mental measurement lines to each other. In so doing, they will develop a better understanding of measurement systems.

Our discussion of instructional studies suggests that the lack of transfer demonstrated by a long history of studies on instruction in measurement estimation, for both guess and check and instruction on using reference points, underscores a serious weakness in these forms of instruction: Individual, disconnected points on a scale may be learned as associations between numbers and referents, but underlying conceptual measurement scales are not constructed. We suggest that without this fundamental level of knowledge, measurement as the driving conceptual framework for measurement estimation will not be understood as a system of interrelated units on scales.

Implications for Research on Measurement Estimation Instruction

Instructional Implications

Our discussion of instructional methods for measurement estimation suggests that there is some basis for using guess and check to teach children about measurement estimation. It could be a conceptually richer form of instruction than appears at first glance, particularly if supplemented along the lines suggested above.

We raise the question of whether learning to use reference points for measurement estimation is conceptually distinct from the guess and check method: reference points may simply serve as mnemonics for associations between specific magnitudes and their corresponding measurements. Although this is an important aspect of number sense, if having a meaningful set of reference points is to lead to a more elaborated understanding of measurement systems, students will likely need instruction on the principles and constraints that govern measurement estimation along the lines of what we suggested above for the guess and check method. Students might also be asked to generate reference points for "benchmark" units, for example, 6 inches, rather than estimating with only single unit reference points such as a paper clip for an inch. Reference points could also be used to help students understand the relations among different units, for example, understanding that three pens (each 6 inches long) are equivalent in measure to one and a half feet. Such activities would help take reference points beyond mnemonics for single units to

tools for understanding scales and systems of measurement.

Methodological Implications

The guess and check method has been shown to be a reliable way of obtaining a limited amount of improvement in estimation accuracy as a result of a relatively short intervention. Because of this robust finding, researchers interested in examining the efficacy of instruction, such as strategies training for measurement estimation, would be well advised to select a method of instruction other than guess and check for the control group. In Hildreth's (1980) dissertation, for example, both the strategies and the guess and check group improved in their estimation accuracy, leading to the finding that there was no significant difference between the two groups at posttest. Had Hildreth pitted his strategies group against another type of control group, he may have found statistical evidence for the efficacy of strategies training. There is a great need for research on instruction in measurement estimation strategies, and perhaps recognition of the problem presented by using guess and check as a control will stimulate the design of studies that are likely to find significant differences between conditions.

Future studies designed to investigate the effects of instructing students on strategies for estimating need to clearly distinguish strategies instruction from guess and check. Previous researchers often embedded guess and check within strategies instruction (Attivo & Trueblood, 1980; Hildreth, 1983; Markovits, 1987), making it impossible to determine whether strategies instruction alone can improve measurement estimation competencies. Because we can be confident that using the guess and check method to teach measurement estimation will result in improvements in estimation performance, witnessing increases in participants' estimation skills when using combination methods does not demonstrate anything new. What is needed is a series of studies that examines and contrasts the use of individual strategies for measurement estimation.

Implications for Research on the Cognitive Processes in Measurement Estimation

Those studying strategies for measurement estimation may find new directions for research from examining findings on computational estimation strategies (e.g., Reys, Rybolt, Bestgen & Wyatt, 1982; Shimizu & Ishida, 1994; Sowder & Wheeler, 1989). Researchers in this area have identified a variety of strategies that are used by good estimators that may also apply to measurement estimation, such as a tolerance for error (Shimizu & Ishida, 1994). Further, estimators may use strategies that operate directly on the estimate itself, not yet identified in measurement estimation, such as the rounding and truncating strategies identified in computational estimation (Sowder & Wheeler, 1989).

The issue of the role of competency in physical measurement for measurement estimation is in dire need of being addressed. Although many researchers who have investigated measurement estimation skills also gave students general tests of mathematical ability (Corle, 1960; Crawford & Zylstra, 1952; Forrester & Shire, 1994; Hildreth, 1980; Immers, 1983; Morgan, 1986), they did not include an assessment of estimators' understanding of physical measurement. Further, studies of measurement estimation do not typically include an investigation of how accurately individuals mentally represent the units they use to estimate. This appears to be a significant gap in the literature because without studies which together assess individuals' understanding of physical measurement, representations of units, and measurement estimation skills, it is impossible to determine where observed difficulties in measurement estimation lie. Such studies would greatly inform the design of interventions on measurement estimation whose purpose is to enhance estimation skills or to serve as a vehicle for promoting understanding of physical measurement.

Although the framework presented in this paper is applied to only linear measurement estimation, it may provide a tool for examining the cognitive processes involved in estimating other attributes such as weight. Research is needed to determine whether different attributes elicit the different kinds of estimation processes that are described in the framework. For example, unit iteration may be commonly used for length and area estimation because it is possible to directly segment an object into units that can be visualized, whereas when estimating weight or temperature it is not possible to do so. Instead, the direct access route described above may lend itself well to weight and temperature. Additional cognitive processes, not yet identified, may be mobilized in the estimation of attributes other than length.

Conclusions

Why Measurement Estimation is Difficult

Gelman and colleague (Gallistel & Gelman, 1992; Gelman, 1991) have proposed that the nonverbal counting system provides a structure for learning counting and the basic arithmetic facts. They have also argued that mathematical knowledge that does not map well to the nonverbal counting system will present conceptual difficulties for young children. Thus, they propose that it is the degree of isomorphism between the mathematics to be learned and the nonverbal counting system that predicts the relative difficulty of learning mathematical concepts for young children. For example, there are many rational numbers between any two natural numbers, unlike natural numbers which are represented only at discrete intervals along the number line. This is counter-intuitive, Gallistel and Gelman (1992) note, because there is no provision for representing points between natural numbers in the preverbal scheme. This is one reason that children find rational numbers difficult to learn (Gelman, 1991).

Measurement estimation is an interesting case because in some ways it maps very well to the nonverbal counting system–the number line representing numerosity can also easily represent measurements, as long as they are expressed as whole numbers[5]. It is getting to and from this number line that we have argued presents difficulties for estimators: in order to verbally count units, the estimator must decide where the divisions between countable entities lie. We have specified above how complex a mental act this is, requiring knowledge of physical measurement concepts, the recollection or construction of an image of a unit, the ability to spatially manipulate this unit, and the need to keep track of units counted while continuing to segment the TBE. Future research is needed to tease apart the relative contributions of these aspects of the measurement estimation process in order to make instructional recommendations for those areas which are most problematic.

Implications for Instruction

We have suggested that key to facility in measurement estimation is the construction of a mental measurement line representing nonverbal magnitudes and their corresponding measurements. An estimator must be able to map from measurements to nonverbal magnitudes, and also from nonverbal magnitudes to measurements through reference to this line. We have suggested that although instruction, using both guess and check and reference points, functions to build these mappings between specific measurements and their nonverbal magnitudes, these resulting associations will be piecemeal and unsystematic. Further, efforts have not moved beyond fostering associations between individual units and their corresponding nonverbal magnitudes–after Resnick (1989), we have called this the *semantics* of measurement.

We propose that the next step in instruction on measurement estimation is to promote the development of a *syntax* of measurement. By this we mean understanding how meaningful units are ordered on scales, and how different scales are related to each other within and across different measurement systems. In summary, we have suggested that educators encourage the development of reference points for a variety of single and multiple units and practice aligning these on mental number lines. Inferences should be encouraged for points along the number lines not represented by reference objects. Movement across different mental measurement lines can be practiced, for example, using the image of a 6-inch long pen to find the 1 ½ foot point on a mental measurement line for feet.

The pragmatic aspect of measurement estimation must also be supported as in Markovits's (1987) intervention–that is, the understanding of situational constraints on measurement. For example, one may have an absolute understanding of the magnitude of 12 inches, but to know that 12 inches is a relatively large measurement in one context but not another requires the development of the understand of relations between situations and measurements.

To fully flesh out our analogy of learning a language and learning measurement estimation, we must add that the analogue to grammatical rules are provided by knowledge of the principles and constraints of physical measurement. We have made the point that instructional efforts have had a focus, albeit a limited one, on the semantics of measurement, and that what is needed to promote facility with measurement estimation is additional attention to the syntax and pragmatics of this mathematical domain.

Finally, the cultural context of measurement estimation has yet to be explored. In seminal work by Gay and Cole (1967), illiterate Kpelles of Liberia were found to estimate more accurately than American Peace Corps volunteers when using the kopi, a pervasive unit in their daily commerce. The findings of Gay and Cole's (1967) research suggests that the function of measurement estimation within a culture, and the extent to which measurement estimation permeates daily life as opposed to being encapsulated in a formal context, will predict how well individuals learn to estimate. It is not surprising that temperature emerged in the studies of measurement estimation performance reviewed above as the attribute individuals were able to estimate most accurately–most people hear reports of temperatures every day, and base decisions about clothing and behavior on these numbers.

These observations remain speculative however; what is needed is systematic research that examines how measurement estimation is used in daily life, by typical individuals as well as tradespeople, and how its cultural role contributes to the development of competence in this domain.

Notes

[1] Moskol (1980) found that 56% of students who imagined a door accurately estimated the door width, whereas 82% of those who looked at a door did so. The relatively poor performance of those who imagined the door may be explained by the variability in the imagined objects.

[2] Percent error is calculated by obtaining the absolute difference between the estimate and the actual measurement of the to-be-estimated item, and dividing this difference by the actual measurement of the item (e.g., an estimate of 6 inches for a 4-inch long pencil would yield a percent error of 50%, obtained through the following calculations: 4 - 6 = -2, 2/4 = .50; .50 = 50%).

[3] Power functions are used in *psychophysics* to examine relationships between perceived magnitudes and estimates (e.g., whether doubling the intensity of light results in the perception that the light is twice as strong), and for comparing judgments across attributes, (e.g., whether doubling sound decibels results in the same degree of increase in estimates as does doubling the intensity of light; Geschieder, 1997). Power functions reveal only the proportional relationship between subjective and actual magnitudes.

[4] We use the term *syntax* here more broadly than Resnick (1989), who refers to syntax in mathematics as rules for manipulating numbers. We base our analogy of syntax in mathematics on a definition which includes an understanding of the "orderly or systematic arrangement of parts or elements; constitution (of body); a connected order or system of things" (Simpson & Weiner, 1989).

[5] Almost all studies we reviewed, with the exception of Siegel et al., (1982), did not require estimators to use fractions of units. We assume that when using fractions of units to estimate measurement, children will additionally have the same kinds of difficulties with fractions that have been documented in the literature (see Gelman, 1991).

References

Attivo, B., & Trueblood, C. R. (April, 1980). *The effects of three instructional strategies on prospective teachers' ability to transfer estimation skills for metric length and area.* Paper presented at the 53rd Annual Meeting of the National Association for Research in Science Teaching, Boston, MA.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International mathematics and science study.* Chesnut Hill, MA: TIMSS International Study Center.

Bell, M. S. (1974). What does "everyman" really need from school mathematics? *Mathematics Teacher, 67,* 196-202.

Boulton-Lewis, G. M. (1987). Recent cognitive theories applied to sequential length measuring knowledge in young children. *British Journal of Educational Psychology, 57,* 330-342.

Bright, G. W. (1976). Estimation as part of learning to measure. In D. Nelson & R. E. Reys (Eds.), *Measurement in School Mathematics: 1976 Yearbook* (pp. 87-104). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Bright, G. W. (1979). Measuring experienced teachers' linear estimation skills at two levels of abstraction. *School of Science and Mathematics, 79,* 161-164.

Campbell, P. F., Fein, G. G., & Schwartz, S. S. (1991). The effects of LOGO experience on first-grade children's ability to estimate distance. *Journal of Educational Computing Research, 7,* 331-349.

Carpenter, T. P., Kepner, H., Corbitt, M. K., Lindquist, M. M., & Reys, R. E. (1980). Results and implications of the Second NAEP Mathematics Assessments: Elementary School. *The Arithmetic Teacher, 27,* 10-12; 44-47.

Carter, H. L. (1986). Linking estimation to psychological variables in the early years. In H. L. Schoen & M. J. Zweng (Eds.), *Estimation and Mental Computation: 1986 Yearbook* (pp. 74-81). Reston, VA: The National Council of Teachers of Mathematics.

Cassell, M., E. (1941). *What measures do children know, and why?* Unpublished Ed. D. Thesis, Boston University, Boston.

Clayton, J. (1988). Estimation. *Mathematics Teaching, 125,* 18-19.

Coburn, T. G., & Shulte, A. P. (1986). Estimation in measurement. In H. L. Schoen & M. J. Zweng (Eds.), *Estimation and Mental Computation: 1986 Yearbook* (pp. 195-203). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Corle, C. G. (1960). A study of the quantitative values of fifth and sixth grade pupils. *The Arithmetic Teacher, 7,* 333-340.

Corle, C. G. (1963). Estimates on quantity by elementary teachers and college juniors. *Arithmetic Teacher, 10,* 347-357.

Crawford, B. M., & Zylstra, E. W. (1952). A study of high school seniors' ability to estimate quantitative measurements. *Journal of Educational Research, 55,* 241-248.

Dahaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1-42.

Davydov, V. V., & Tsvetkovich, Z. H. (1991). On the objective origin of the concept of fractions. *Focus on Learning Problems in Mathematics, 13,* 13-65.

Duncan, D. R., & Litwiller, B. H. (1986). Benchmark: Two folk methods for estimating distance. In H. L. Schoen & M. J. Zweng (Eds.), *Estimation and mental computation* (pp. 223-224). Reston, VA: National Council of Teachers of Mathematics.

Ferris, S. H. (1972). Improvement of absolute distance estimation underwater. *Perceptual and Motor Skills, 35,* 299-305.

Forrester, M. A., Latham, J., & Shire, B. (1990). Exploring estimation in young primary school children. *Educational Psychology, 10,* 283-300.

Forrester, M. A., & Shire, B. (1994). The influence of object size, dimension, and prior context on children's estimation abilities. *Educational Psychology, 14,* 451-465.

Friebel, A. C. (1967). Measurement understandings in modern school mathematics. *Arithmetic Teacher, 14,* 476-480.

Gallistel, C. R., & Gelman, R. (1991). Subitizing: The preverbal counting process. In W. E. Kessen, A. Ortony, & F. I. M. Craik (Eds.), *Thoughts, memories, and emotions: Essays in honor of George Mandler* (pp. 65-81). Hillsdale, NJ: Erlbaum.

Gallistel, R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43-74.

Gay, J., & Cole, M. (1967). *The new mathematics and an old culture: A study of learning among the Kpelle of Liberia.* New York: Holt, Rinehart & Winston.

Gelman, R. (1991). Epigenetic foundations of knowledge structures: Initial and transcendent constructions. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 293-322). Hillsdale: Erlbaum.

Gescheider, G. A. (1997). *Psychophysics: The fundamentals.* (3rd ed.). Hillsdale, NJ: Erlbaum.

Gibson, E. J., & Bergman, R. (1954). The effect of training on absolute estimation of

PP1 (11).max

distance over the ground. *Journal of Experimental Psychology, 48*, 473-482.

Gibson, E. J., Bergman, R., & Purdy, J. (1955). The effect of prior training with a scale of distance on absolute and relative judgments of distance over the ground. *Journal of Experimental Psychology, 50*, 97-105.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25-49). Cambridge, MA: Massachusetts Institute of Technology.

Hancock, S. J. C. (1994). *The mathematics and mathematical thinking of seamstresses.* Unpublished Doctoral Dissertation, The University of Georgia, Athens, G.A.

Hartley, A. A. (1977). Mental measurement in the magnitude estimation of length. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 622-628.

Hartley, A. A. (1981). Mental measurement of line length: The role of the standard. *Journal of Experimental Psychology, 7*, 309-317.

Hiebert, J. (1981). Units of measure: Results and implications from national assessment. *Arithmetic Teacher, 28*, 38-43.

Hiebert, J. (1984). Why do some children have trouble learning measurement concepts? *Arithmetic Teacher, 31*, 19-24.

Hildreth, D., J. (1980). *Estimation strategy uses in length and area measurement tasks by fifth and seventh grade students.* Unpublished Doctoral Dissertation, The Ohio State University, Ohio.

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

Immers, R., C. (1983). *Linear estimation ability and strategy use by students in grades two through five.* Unpublished Doctoral Dissertation, University of Michigan, Ann Arbor, M.I.

Jones, M. L., & Rowsey, R. E. (1990). The effects of immediate achievement and retention of middle school students involved in a metric unit designed to promote the development of estimating skills. *Journal of Research in Science Teaching, 27*, 901-913.

Joram, E., Gabriele, A. J., Gelman, R., & Subrahmanyam, K. (April, 1996). *Building meaning for units of measurement: A personal anchors approach.* Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

King, I., & Whitman, N. (1973). Going metric in Hawaii. *Arithmetic Teacher, 20*, 258-260.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 47-60.

Kosslyn, S. M., Margolis, J. A., Barrett, A. M., Goldknopf, E. J., & Daly, P. F. (1990). Age differences in imagery abilities. *Child Development, 61*, 995-1010.

Kosslyn, S. M., Pick, H. L., Jr., & Fariello, G. R. (1974). Cognitive maps in children and men. *Child Development, 45*, 707-716.

Kosslyn, S. M., Reiser, B. J., Farah, M. J., & Fliegel, S. L. (1983). Generating visual images: Units and relations. *Journal of Experimental Psychology: General, 112*, 278-303.

Kouba, V. L., Brown, C. A., Carpenter, T. P., Lindquist, M. M., Silver, E. A., & Swafford, J. O. (1988). Results of the Fourth NAEP Assessment of Mathematics: Measurement, geometry, data interpretation, attitudes, and other topics. *Arithmetic Teacher, 35(9)*, 10-16.

Lea, G. (1975). Chronometric analysis of the method of loci. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 95-104.

Levin, J. A. (1981). Estimation techniques for arithmetic: Everyday math and mathematics instruction. *Education Studies in Mathematics, 12*, 421-434.

Luria, S. M., Kinney, J. S., & Weissman, S. (1967). Distance estimates with "filled" and "unfilled" space. *Perceptual and Motor Skills, 24*, 1007-1010.

MacNeal, E. (1994). *Mathsemantics: Making numbers talk sense.* New York: Penguin Books.

Markovits, Z. (1987). *Estimation: Research and curriculum development.* Unpublished Doctoral Dissertation, Weizmann Institute of Science, Rehovot, Israel.

Markovits, Z., Hershkowitz, R., & Bruckheimer, M. (1989). Number sense and nonsense. *Arithmetic Teacher, 36*, 53-55.

Masingila, J. O. (1993). Learning from mathematics practice in out-of-school situations. *For the Learning of Mathematics, 13*, 18-22.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*, 320-334.

Morgan, V. R. L. (1986). *A comparison of an instructional strategy oriented toward mathematical computer simulations to the traditional teacher-directed instruction on measurement estimation.* Unpublished Doctoral Dissertation, Boston University, Boston, M.A.

Moskol, A. E. (1980). *An exploratory study of the processes that college mathematics students use to solve real-world problems.* Unpublished Doctoral Dissertation, University of Maryland, Maryland.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Neufeld, K. A. (1989). Body Measurement. *The Arithmetic Teacher, 36*, 12-15.

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

Paulos, J. A. (1988). *Innumeracy - mathematical illiteracy and its consequences.* New York: Hill and Wang.

Piaget, J., Inhelder, B., & Szeminska, A. (1960). *The child's conception of geometry.* New York: Basic Books.

Poynter, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition, 11*, 77-82.

Pressey, A. W. (1974). Age changes in the Ponzo and filled-space illusions. *Perception & Psychophysics, 15*, 315-319.

Resnick, L. B. (1983). A developmental theory of number understanding. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 109-151). New York: Academic Press.

Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist, 44*, 162-169.

Reys, R. E., Rybolt, J. E., Bestgen, B. J., & Wyatt, J. W. (1982). Processes used by good computational estimators. *Journal for Research in Mathematics Education, 13*, 183-201.

Rowsey, R. E., & Henry, L. L. (1978). A study of selected science and mathematics teacher education majors to assess knowledge of metric measures and their applications. *Journal of Research in Science Teaching, 15*, 85-89.

Scott Foresman/Addison Wesley (1998). *Scott Foresman - Addison Wesley Math.* Menlo Park, CA: Addison Wesley Longman.

Shaughnessy, C. A., Nelson, J. E., & Norris, N. A. (1997). *NAEP 1996 Mathematics cross-state data compendium for Grade 4 and Grade 8 assessment.* Technical Report. Washington, DC: National Center for Education Statistics.

Shimizu, K., & Ishida, J. (1994). The cognitive processes and use of strategies of good Japanese estimators. In R. E. Reys & N. Nohda (Eds.), *Computational alternatives for the twenty-first century: Cross-cultural perspectives from Japan and the United States* (pp. 161-178). Reston, VA: National Council of Teachers of Mathematics.

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*, 211-232.

Simpson, J. A., & Weiner, E. S. C. (1989). *The Oxford English dictionary.* (Vol. XVII). Oxford: Clarendon Press.

Smith, C. R. (1994). *Learning disabilities: The interaction of learner, task, and setting.* (3rd ed.). Boston: Allyn and Bacon.

Sowder, J. (1992). Estimation and number sense. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 371-389). New York: Macmillan Publishing Company.
Sowder, J. T., & Wheeler, M. M. (1989). The development of concepts and strategies used in computational estimation. *Journal for Research in Mathematics Education, 20*, 130-146.
Spitzer, H. F. (1961). *The teaching of arithmetic*. Boston: Houghton Mifflin Company.
Swan, M., & Jones, O. (1971). Distance, weight, height, area and temperature percepts of university students. *Science Education, 55*, 353-360.
Swan, M., & Jones, O. (1980). Comparison of students' percepts of distance, weight, height, area, and temperature. *Science Education, 64*, 297-307.
Thompson, C. S., & Rathmell, E. C. (1989). By way of introduction. *Arithmetic Teacher, 36*, 2-3.
Thorndike, E. L. (1927). The law of effect. *American Journal of Psychology, 39*, 212-222.
Thorndyke, P. W. (1981). Distance estimation from cognitive maps. *Cognitive Psychology, 13*, 526-550.
Trowbridge, M. H., & Cason, H. (1932). An experimental study of Thorndike's theory of learning. *Journal of General Psychology, 7*, 245-258.
Usiskin, Z. (1986). Reasons for estimating. In H. L. Schoen & M. J. Zweng (Eds.), *Estimation and mental computation* (pp. 1-15). Reston, VA: National Council of Teachers of Mathematics.
Van de Walle, J., & Thompson, C. S. (1985). Estimate how much. *Arithmetic Teacher, 35*, 4-8.
Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*, 130-137.
Wilson, D. W. (1936). *What measures do people know, and why*. Unpublished Master's Thesis, Boston University, Boston.
Wilson, G. M., & Cassell, M. (1953). A research on weights and measures. *Journal of Educational Research, 46*, 575-585.

APPENDIX A
Summary of reviewed education studies on performance

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Task description	Estimation (how measured)	Scoring	Outcomes
Cassell (1941) Wilson & Cassell (1953)	6573 2nd- through 10th-grade students	Height, width Distance	Customary	Items absent Units absent	4 free response items	"Correct" and "reasonable" estimates specified for each item. Percent error permitted for "correct" responses varied by item.	1. 18% of students on average at each grade level gave correct responses. 2. Accuracy increased with grade level. 3. Adults scored 19-56% correct across items.
Clayton (1988)	1000 British children spanning the "entire age and attainment range"	Unspecified; Given example of length item	Metric, Nonstandard	Items present/absent Units present/absent	Unspecified number of free-response items	Reports how many children give estimates within ±20% the true measurement.	1. Most children respond within 20% of true measurement, but only when answer is less than 100 (Frequencies not provided).
Corle (1960)	147 5th- and 6th-grade students	Length, circumference (also weight, time, temperature, liquid capacity)	Customary	Items present Units absent	10 free response items	Percent error. Responses scored as 100% error included estimates exceeding 100% error, estimates not attempted, and estimates made using incorrect terminology.	1. For estimates of linear measurements, when outliers are excluded average percent error for 5th-graders is 96% and 56% for 6th-graders.
Corle (1963)	1. 96 college juniors 2. 368 elementary teachers	Length, circumference (also weight, time, temperature, liquid capacity, volume, barometric pressure)	Customary	Items present Units absent	10 free response items	Percent error. Also classified scores as: correct, within 10%, over 10% and under 100%, over 100% error	1. Average percent error was 21-53% for estimates of linear measurements. 2. No differences found between teachers and undergraduates. 3. Majority of estimates over 10% and under 100% error.

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Task description	Estimation (how measured)	Scoring	Outcomes
Crawford & Zylstra (1952)	102 high-school seniors	Not specified (Examples show questions on height and length).	Customary	Items present/absent Units absent	32 multiple choice items	Choices were classified as "exact correct" or "nearly correct" (approximately 20-25% error permitted). In example, 3 of 5 choices were "correct" or "nearly correct".	1. Students obtained an average of 42% correct. 2. Average "nearly correct" not reported. On one example given, it is approximately 80%. 3. Presence/absence of objects did not affect estimation accuracy.
Forrester, Latham, & Shire (1990)	70 5- to 8-year-old children	Length (also area, volume)	Nonstandard	Items present Units present	6 free response items. Retrospective probing for methods used to make estimates	Analyses performed on log-transformed scores. Classified methods used to make estimates.	1. Children underestimated on length items. 2. Accuracy increased with age.
Forrester & Shire (1994)	67 8- to 11-year-old children	Length (also area, volume)	Nonstandard	Items present Units present	10 free response items. Retrospective probing for methods used to make estimates	Percent error scores. Classified methods used to estimate.	1. Average percent error on length items was 10% for 8-9 year-olds and 7% for 10-11 year-olds. 2. Younger children underestimated more frequently than older children, but only on relatively long items (16 units vs. 4 and 8). 3. Underestimates tended to provide "immediate responses" in contrast to good estimators or overestimators.

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Task description	Estimation (how measured)	Scoring	Outcomes
Mosbot (1980)	30 undergraduate mathematics majors	Length	Customary	Items present/absent Units absent	Participants thought aloud as they solved 2 estimation problems.	Strategies were classified. Estimates considered "correct" if within ±10% of actual measurement of TBE.	1. 83% correct for estimates of door width but only 7% correct for estimates of width of highway lane. 2. Inaccurate comparisons were associated with incorrect estimates; accurate comparisons and the use of an imaginary yardstick were associated with correct estimates.
Rowney & Henry (1978)	32 preservice teachers majoring in mathematics or science	Length, area, volume, temperature	Metric	Items present Units absent	20 free response estimation items. Participants also completed a 20 item test requiring them to convert among metric units.	A panel of judges determined a range of "correct" responses for each item. Information about criteria not provided. Outcomes not presented nor analyzed as function of attribute.	1. Accurate at converting one metric unit to another (average of 77% correct). 2. Scored much lower on estimating test (average of 43% correct).
Siegel, Goldsmith & Madson (1982)	1. 140 students in grades 2 through 8 2. 10 adults	Length, height	Standard	Items present Units present	24 free response items. Probing for strategy use	Estimates classified as: "reasonable" (≈100% error) and "accurate" (≤50% error). Classified responses into strategies	1. Objects requiring application of a single unit were estimated more accurately than those requiring fractional or multiple units. 2. Least accurate estimates were made for objects requiring mental decomposition prior to estimating. 3. Accuracy increased with age for all but those items lending themselves to a regular decomposition strategy.

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Task description	Estimation (how measured)	Scoring	Outcomes
Swan & Jones (1971)	392 preservice teachers	Length, distance, (also weight, area, and metric temperature).	Customary and metric	Not specified whether items were present. Units absent	13 free response items	Responses classified as "reasonable" if within 25% of TBE	1. For distance, 20% of estimates were "reasonable" in metric; 47% in customary. For height, 64% of estimates were reasonable in customary (metric not reported for height). 2. Estimates in metric units were less accurate than those made in customary units.
Swan & Jones (1980)	1. 304 elementary; 266 junior-high, and 234 high-school students 2. 554 adults	Length, distance, (also weight, area, and temperature)	Customary and metric	Items present. Units absent	15 free response items	Responses classified as "reasonable" if within 25% of TBE.	1. For length and distance, 32-57% of estimates were accurate in customary units; 29-42% in metric units.
Wilson (1936)	2056 3rd- through 12th-grade students	Height, width	Standard	Items absent. Units absent	5 free response items	Responses classified as "fair estimates" or "wrong" (criteria for classification not given). An example given showed fair estimates ranged from 28 to 34 inches for a 30 inch TBE.	1. Reasonable estimates increased with grade from an average of 3% of students in 3rd grade to an average of 31% in 12th grade.

APPENDIX B

Summary of reviewed instructional studies on measurement estimation

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Treatment description	Outcomes (pretest/posttest)	Responses (how scored)	Outcomes
<u>Guess and Check Instruction.</u>							
Perris (1972) Study I	40 Navy-enlisted men	Distance	Customary	Participants either: made 15 estimates underwater with feedback after each; made 15 estimates underwater with no feedback; played checkers underwater; or swam underwater	10 items administered twice at pre- and posttests	Analyses performed on estimates and on log-transformed scores.	1. Estimation accuracy of feedback group improved more than the other three groups.
Study II	20 Navy-enlisted men	Distance	Customary	All participants trained in air, using guess and check.	10 items, each administered once in air and once in water at both pre- and posttest. For half participants, water was clear, for other half it was turbid at pre- and posttesting.	Analyses performed on estimates and on log-transformed scores.	1. All participants improved their estimates in air. 2. The clear water group improved considerably but the turbid water group did not.
Study III	12 Navy-enlisted men	Distance	Customary	All participants trained in moderately clear water using guess and check	Given pretest, first and third posttests in moderately clear water. Half participants posttested second time in turbid than clear water; the other half in clear than turbid water	Analyses performed on estimates and on log-transformed scores.	1. Estimates were less accurate in clear water and more accurate in turbid water than at pretest.
Gibson & Bergman (1954)	1. 32 undergraduates 2. 92 airmen in training	Distance	Customary	Experimental group made 90 estimates of distances of targets 39 to 435 yards away; each estimate immediately corrected by experimenter. Control group completed an aptitude test.	18 item pre- and posttest	Percent error	1. Estimation accuracy of treatment group improved more from pre- to posttest than control group.

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Treatment description	Outcomes (pretest/posttest)	Responses (how scored)	Outcomes
Gibson, Bergman & Purdy (1955)	70 airmen in training	Distance	Customary units for pre- and posttesting Nonstandard units for training	Experimental group made 10 distance estimates by blowing a whistle to indicate where a bicycle rider should stop. The rider then placed a marker at the correct spot, showing the estimator their degree of error. Control group pre- and posttested only.	18 item pre- and posttest. 4 additional estimates made at posttest.	Estimates converted to logarithms, and a "constant error of estimation obtained for each subject"	1. Estimation accuracy of treatment group greater at posttest than control group.
Immers (1983)	48 students in 2nd through 5th grade	Length	Nonstandard	Treatment group received 4 lessons in which they guessed and checked line lengths (1 to 4 in.), followed by 6 practice sessions in which they guessed and checked longer line lengths (1 to 8 in.) Control group received no instruction	Estimated lengths of 24 lines and drew lengths of 4 lines on pre- and posttests. Interviewed 8 students from the experimental group to see what strategies they used.	Percent error classified into 7 pt. scale, with "0" meaning >30% error and "7" meaning no error. Classified responses into "correct" (<30% error) and "incorrect" (>30% error)	1. Treatment group more accurate than control only when posttested with same kinds of items used in training. 2. Strategies of treatment group stayed the same from pre- to posttesting. 3. Accuracy increased with grade level on posttest.
Jones & Rowsey (1990)	397 7th-grade students	Length (also mass, liquid capacity)	Metric	Treatment group estimated then measured (guess and check) over five lessons. Control group measured without estimating over five lessons.	Students given pre-, post, and delayed posttest (five weeks after instruction) with same 20 item test.	Mean score correct reported. Scoring criteria not provided.	1. No differences found on immediate posttest performance between treatment and control groups. 2. Treatment group scored higher than control group on delayed posttest. 3. Authors report that mean scores of both groups on all tests were low.

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Treatment description	Outcomes (pretest/posttest)	Responses (how scored)	Outcomes
Morgan (1986)	46 4th-grade students	Length (also area)	Customary	Half students estimated objects on a computer, half worked with paper and pencil. Both groups received 90 minutes of guess and check instruction per week for 4 weeks.	Pre- and posttested on estimation of length, length using a scale, and [illegible]	Scores were absolute values of the difference between estimates and actual measurements of TBEs.	1. No difference in estimation accuracy between computer and paper and pencil groups on posttest. 2. On posttest task identical to that used in training, both groups improved dramatically from pre- to posttest.
Thorndike (1927) Study I	11 unspecified	Length	Metric	Treatment group estimated lengths of 3 to 27 cm. long paper strips and heard "right" or "wrong" after each estimate. Control group estimated with no feedback.	Pre- and posttested on 50 paper strips ranging from 3 to 27 cm. in lengths.	Deviations from true lengths.	1. Feedback group improved more than control group in accuracy of estimates from pre- to posttest.
Study II	30 adults	Length	Customary	Drew 5400 3- to 6-inch long lines while blindfolded over a 7-day period. Heard "right" if line was acceptable range; "wrong" if not. A control group drew lines but received no feedback.	Pre- and posttested on line drawing (given a measurement, they drew a line of corresponding length)	Drawings were "right" when within 1/8th inch of a 3- inch long line, or within 1/4 inch of the other lengths. Analyzed percent "right."	1. Feedback group improved more than control in accuracy of line drawing from pre- to posttest.
Trowbridge & Cason (1932)	60 university students	Length	Customary	Four groups drew 3- to 6- inch long lines while blindfolded over 100 trials in one of four conditions: No feedback given; the experimenter stated a nonsense syllable after each line was drawn; participants told "right" or "wrong"; or participants told how many eighths of an inch error they had made.	No pre- or posttest given. Compared successive groups of 10 trials.	Drawings were "right" when within 1/8th of 3-inch long or 1/4 inch of other lengths. Analyzed percent "right."	1. Group told exact amount of deviation from correct line length improved more than group given verbal feedback.

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Treatment description	Outcomes (pretest/posttest)	Responses (how scored)	Outcomes
Strategies Instruction							
Attivo & Trueblood (1980)	126 preservice teachers	Length (also area)	Metric	Teachers either: stated personal reference measure, estimate, and checked by measuring; cut or drew a reference unit, checked accuracy of unit, placed it out of sight, guessed and checked; or guessed and checked.	Pre- and posttested on 60 items, including volume.	Mastery defined as with 15% of actual measurement of TBEs on at least 80% of items. Analyses performed on numbers of students in each group achieving mastery. Also compared pre- and posttest error scores.	1. When estimating same materials as in training students who stated reference point performed better than those who drew and removed it. 2. No differences found between reference point and guess and check groups. 3. When estimating with materials that differed from training, no differences found between the groups; all performed poorly.
Hildreth (1980; 1983)	1. 48 5th- and 7th-grade students 2. 24 college freshmen	Length, (also area, perimeter)	Metric for instruction. Metric and customary for pre- and posttests	Two groups received five lessons on estimation: Strategies group were instructed on estimation strategies. Checked measurement after each estimate. Second group instructed on guess and check.	Pre- and posttested on 24 item test. All students interviewed and both estimate and strategy recorded.	"Correct" response defined as within one-third of measurement of TBE. Strategies were classified and tabulated.	1. Strategies group used greater number of strategies at posttest than guess and check group. 2. No differences found in estimation performance at posttest. 3. No differences found in estimation ability as a function of grade level.
Joram, Gabriele, Gelman & Subrahmanyam (1996)	36 4th-grade students	Length	Customary	Treatment group given 5 lessons in which they practiced visualizing, drawing, and estimating with a personal reference point. Students shown how to iterate reference point. Control group practiced guessing and checking in 5 sessions.	All students pre- and posttested: 1. Drew an inch and a foot 2. Estimated lengths of 16 items, half ropes, half other objects.	Percent error calculated and contrasted across the two groups for drawings and estimates.	1. Treatment group's drawings more accurate than control group at posttest. 2. Treatment group improved more than control group but only when estimating rope in inches. Both groups improved estimation accuracy in feet.

Author (date)	Participants (academic level)	Attribute(s)	Unit(s)	Treatment description	Outcomes (pretest/posttest)	Responses (how scored)	Outcomes
Other Instructional Studies							
Bright (1979)	21 middle school teachers	Length	Nonstandard	All participants enrolled in summer workshop on the metric system (description not provided). 10 guess and check practice items given before test.	Test performed on day 3 of workshop.	Percent error	1. Reports that skills of teachers improved with practice.
Campbell, Fein, & Schwartz (1991)	48 1st-grade students	Distance	Nonstandard	Treatment group received Logo instruction 1 hr. per week for 6 months. Children were not asked to estimate distance during lessons but focused on geometric concepts and motion. Control group received no instruction.	Posttested on 24 items in two sessions. Test items were presented on a computer screen.	Analyzed ratio of estimates to actual distance.	1. Children who had Logo instruction were more accurate at estimating distance. 2. Longer lines resulted in less accurate estimates.
Friebel (1967)	185 7th-grade students	Distance (also area, volume, weight)	Customary	Treatment group instructed for one year using the <u>School Mathematics Study</u> (information not provided about curriculum). Control group received traditional instruction.	All students given written posttest on measurement. A subset of 60 students posttested on 10 items. They were asked to estimate, then measure, and to think aloud throughout.	Verbal responses categorized into estimation strategies.	1. Treatment group outscored control group on written test of measurement. 2. No differences on use of strategies found between treatment and control groups.
Markovits (1987)	1. 328 6th-grade and 7th-grade students. 2. 48 preservice teachers. 3. 39 middle school teachers.	Length, distance (also weight, liquid capacity, area, speed)	Metric	Treatment groups received 8-10 lessons on long number sense curriculum. Control groups did not receive instruction.	Treatment group pre- and posttested. Control group posttested only with 21 item test on multiple attributes.	Items were free response and multiple choice. For multiple choice, participants were asked to explain their choice. Reports percent who selected correct choice and also percent who gave correct explanations.	1. Treatment group outperformed control group on understanding of absolute and relative error and reasonable answers. 2. Treatment group did not show clear superiority in estimate accuracy.