Proceedings of the Workshop "An Interdisciplinary Approach to Understanding and Processing Sketch Maps" at the 2011 Conference on Spatial Information Theory (COSIT 11)
September 2011
Wang, J., Broelemann, K., Chipofya, M., Schwering, A., Wallgrün, J.-O. (eds.)

The Use of Sketch Maps as a Basis for Measures of Spatial Knowledge[1]

Kris Lohmann
Department for Informatics
University of Hamburg
Vogt-Kölln-Straße 30
22527 Hamburg

Abstract. We conducted an experiment to compare the effect of different verbal assistance conditions that were designed to support tactile map reading with the aim of enabling the user of the map to build up survey knowledge. One of the tests of spatial knowledge that was used in this experiment was the production of a sketch map. Another test was to answer to verbal questions about coarse-grained relations between objects on the map. The validity of sketch maps and other assessment methods of spatial knowledge is an issue for every researcher concerned with the study of the efficiency of maps, geographic information systems, or other external spatial representations. In this paper, the results of the sketch-mapping task are compared to the results of the answers to the relation questions. This comparison is performed regarding both averaged data and individual relations (data points) between objects. The results support the validity of sketch maps as a basis for measures of the accuracy of spatial knowledge.

Keywords. Sketch maps, validity, methods to test spatial knowledge, cognitive maps, spatial mental models, spatial cognition

1. Introduction

We make use of external spatial representations in every-day scenarios: Finding the way to the next bakery or navigating in a city being new to us. Frequently, we use external representations of space, for example, maps, to gain an overview of a new environment. By interacting with these external representations, we build up an internal representation of space. This representation is often called 'cognitive map', a term introduced by Tolman [1, see also 13]. The term 'cognitive map' has been the subject of discussion and has often been challenged for its implication to be image-like. Tversky [2] provides an overview of cognitive-map research and proposes to rather use the term 'spatial mental model', which we will adopt in this paper. A spatial mental model consists of fragments of spatial information of different kinds that are integrated to various degrees. Hegarty [3] describes the process of building up a spatial mental model as follows: sensory inputs are encoded to construct an internal representation of

[1] The research reported in this paper has been partially supported by DFG (German Science Foundation) in IRTG 1247 'Cross-modal Interaction in Natural and Artificial Cognitive Systems' (CINACS). I thank the anonymous reviewers for their highly useful commentaries.

the environment. After a readout process, this internal representation can be subject of various outcome measures assessing the performance. Often, sketch maps created by the participants are the basis for such an outcome measure. For example, the degree to which they resemble the original map can be estimated by one or more raters.

Methods to test the quality of spatial mental models–humans' internal representations of space–are a frequent issue in conducting research concerning the effectiveness of external representations of space, such as state-of-the-art geographic information systems, conventional visual maps, or tactile maps.

Our research on external representations focuses on tactile maps. As visual maps are not accessible for blind and visually impaired people, tactile maps can be used as substituting external spatial representations. The *Verbally Assisting Virtual-Environment Tactile Map (VAVETaM)* we approach gives situated verbal assistance during exploration. The maps are presented by using advanced human-computer interfaces, particularly the Sensable Phantom Omni (see Figure 1(a)). This interface enables a virtual haptic perception of three-dimensional objects. In the VAVETaM scenario, these objects are virtual counterparts to traditional tactile maps. The maps are realized as plane objects with cavities that, for example, represent streets and buildings (see [4-6], for details). A visualization of a cross-section through such a map is shown in Figure 1(b).

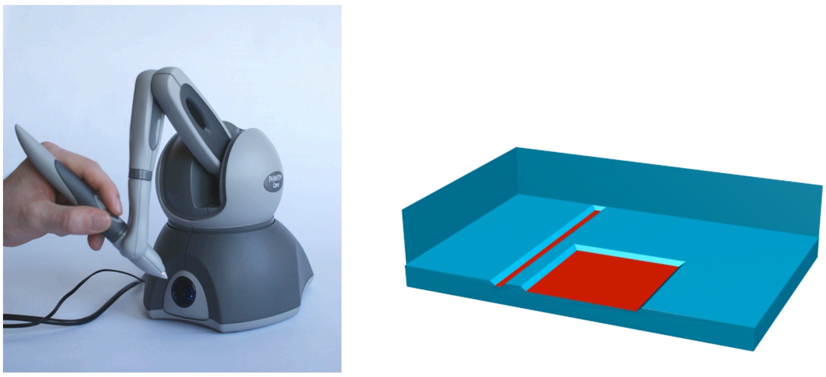

Figure 1. (a) The Sensable Phantom Omni interface used in the experiment. (b) Visualization of a cross-section through a tactile map.

To test the usefulness of the approach, we conducted an experiment. In this experiment, we measured the effect of two different assistance conditions on the spatial mental model a blindfolded user of the map acquired from learning the maps haptically. One verbal-assistance condition called *simple assistance* contained only information about the names of the objects on the map and the other condition called *extended assistance* included additional statements about the spatial layout of the map objects and references to objects without a proper name (e.g., 'the crossing between Bergstraße and Hochstraße').

To investigate the spatial mental model the participants of the experiment build up during exploration, we used mutually supportive tests [for this approach, see also 7]. One of these tests was to draw a sketch map reflecting the spatial layout of the tactile map that was explored in the learning condition. The validity of sketch maps is often questioned in cognitive science: how well do they, as an externalization of an internal representation undergoing readout and transformation processes, reflect humans' internal representations? To which degree do they depend on sketching ability or graphic modeling ability [8]? How do they perform compared to other assessment methods of space?

In the next section, we will provide a review of the literature concerning testing methods of spatial mental models. In this review, the focus will be on the discussion of the validity and reliability of sketch maps. In Sect. 3, the methods and results of comparing the sketch maps produced in the experiment we accomplished to another outcome variable are discussed.

2. Related Work on Testing Spatial Knowledge

A multitude of different means to study spatial knowledge are used in experimental spatial-cognition research, ranging from direction and distance estimates over reconstruction and sketch mapping with varying degrees of parts of the map to be reconstructed or sketched predetermined. Newcombe [8] provides an overview of the different methods used for the study of spatial knowledge and concludes that all methods have different strengths and drawbacks. The author discusses the problematic fact that different methods frequently support differing results. Therefore, more research is needed to compare different methods on both the level of individual data points and the aggregate level.

May [9] provides a taxonomy of the methods used. As the author discusses, spatial mental models can either be externalized by (a) linguistic modeling or (b) graphical modeling. Besides these two general classes, three different classes of methods are used: (1) estimation-based methods, (2) reconstruction-based methods, and (3) chronometric methods.

Estimation-based methods rely on the task of estimating either distances between objects in the spatial layout or angles between objects. Distance-based methods can be divided in methods with absolute judgments, in which participants answer to questions such as 'How far is A away from B?', and methods with relative judgments, in which participants are given a reference distance or a reference angle, for example, 'If A is 100 meters away from B, how far is the distance between C and D?' A third group of tests is based on ordinal judgments. In these tests objects located in space are compared either pairwise or by setting up ranks.

Reconstruction-based methods rely on the task to physically reproduce a model of the original spatial layout. These methods can be classified by the degree to which the model reflects the original layout in respect to the scale, the amount of detail and the dimensionality of the model and the original space. A very high degree of reflection is given if a three-dimensional original space is modeled as a three-dimensional room of equal space. A small degree of reflection of the model to the original space is given when the model is, in comparison with the original space, resized true to scale and realized in two dimensions and incorporation a smaller amount of details than the original.

Chronometric methods are based on theoretic assumptions about underlying mental processes. The latency times of the participants of the experiment for different tasks are measured as outcome variables.

May also discusses sketch maps as reconstruction-based methods, which are frequently used in different contexts following the classical work of Lynch [10]. A major advantage of sketches is their flexibility. For example, sketches can be used to depict a room in a building from a given perspective or to depict survey-type spatial knowledge by sketching a map. However, sketch maps as testing methods are often seen as problematic, as their quality depends on the sketching skills of the sketching participant [11, see also 12].

Kitchin and Blades [13] also provide a classification of methods in use. They distinguish four basic types by the nature of their outcome: (1) uni-dimensional data generation, (2) two-dimensional data generation, (3) recognition tasks, and (4) qualitative approaches. The first type includes all methods that result in a single value, such as distance and direction estimation tasks[2]. Two-dimensional data generation refers to all methods that "elicit a participant's knowledge of configurational spatial relationships […]" [13]. This type of methods includes sketch maps and reconstruction-based methods. The sketch-map methods are further divided into the subclasses (a) basic sketch mapping, (b) normal sketch mapping, (c) cued sketch mapping, (d) longitudinal sketch mapping, and (e) sketch mapping using a sketch mapping language. Methods (a) to (c) differ in how constrained the participants are in drawing the maps. In basic sketch mapping, no further constraints are set up by either the instruction or by the testing material, whereas in normal sketch mapping the experimenter gives more information about what to include in the sketch map, e.g., the participant is asked to label the buildings he knows the name of. Cued sketch mapping is more constrained than normal sketch mapping as the participant is not only asked to freely sketch a map but rather to complete a given portion of a map with specific features. All of these techniques can be combined with the longitudinal sketch mapping approach, which basically records the evolution of the sketch map, for example, by videotaping the sketching process. The last method mentioned above, using a specified sketch map language as suggested by Wood and Beck [e.g., 14], is a technique that is rarely used.

Montello [15] provides an exhaustive overview of methods that can be used to measure mental representations of large-scale distances. The review also includes a discussion of sketch mapping as data-collection method compared to other methods. The author concludes that mapping is very useful for the assessment of the accuracy of configurational spatial knowledge (with some restrictions on the conclusions that should be drawn from the data, see [15] for a discussion).

Blades [16] investigated the reliability of sketch maps by asking participants to draw a sketch map of a familiar route and then draw the same route a week later. If sketch maps are reliable measures of spatial knowledge, the map pairs should be similar. In fact, Blades' study indicated that the evaluation of sketch maps is a reliable method.

Billinghurst and Weghorst [17] studied the validity of sketch maps as measures for spatial mental models of virtual environments by calculating the correlation of the goodness of sketch maps with self-reported feeling of orientation. They concluded that sketch mapping, as a measure of spatial mental models, is most useful in dense worlds containing a large number of objects.

---

[2] What Kitchin and Blades call 'direction tasks' is similar to what May calls 'Winkeleinschätzungen' [angular judgments].

Due to the fact that all of the methods are indirect methods of the human internal spatial representation with different strengths and drawbacks, Jacobson [7] and May [9] both suggest the application of different, mutually supportive methods.

In this paper, a comparison of two methods is discussed: a) verbal relation questions asking for coarse-grained knowledge and b) rater-evaluated sketch maps. Both methods were designed for the evaluation of the accuracy of survey knowledge gained from a tactile map. Survey knowledge can be characterized as map-like overview knowledge about a spatial layout [cf. 18, 19]. The two tasks we compare request very different abilities of the participant. The production of a good sketch map demands for sketching ability and a spatial mental model that is coherent and integrated to a high degree. Furthermore, the way the task was designed left a lot of decisions to the producer of the sketch maps, for example, whether two parallel separate lines or single lines represent streets. In contrast, answering to the relation questions is a comparatively passive task constrained to a high degree.

3. Method

As mentioned in the introduction, the experiment was conducted in the context of the development of the VAVETaM system. The main purpose of this experiment was the evaluation of VAVETaM. Participants were asked to learn a tactile map presented using the Sensable Phantom Omni human-computer interface under different verbal assistance conditions. A detailed discussion of the main results is in preparation.

The results of the experiment offer the option to compare the results of a sketching task to the results of a verbal task. A method for comparison not only on the aggregate level, but also on the level of single data points will be discussed in this section.

3.1. Participants

Twenty-four blindfolded sighted subjects initially participated in the experiment. Two subjects had to be excluded due to technical problems. One subject was not able to pass the test that concluded the training procedure and was therefore also excluded. Data of three additional subjects was collected, leading to a total of 24 evaluated participants (14 males, mean age: 24.7 years, *SD:* 3.3 years). All participants were compensated by partial course credit or on a monetary basis. They were naïve about the purpose of the experiment. All participants gave written informed consent and reported to speak German on a native-speaker level. All participants used their self-reported primary hand for the exploration of the virtual tactile map.

3.2. Method: Outcome Variables

After learning the virtual tactile map, participants were asked to do three tests: 1) answering to questions assessing relations between different map objects, 2) perform a sketching task, and 3) perform a puzzle-like recognition task. The results of and details on the puzzle-like recognition task are not reported here, as this task does not offer the possibility of a direct comparison with sketch maps.

3.2.1. Relation Questions

After learning the map for eight minutes, the participants solved different tasks. The first task consisted of answering to 20 questions about spatial relations between objects. Where it was possible, those questions asked for spatial relations that were not explicitly stated in the prerecorded assisting utterances for the extended-assistance condition. The experimenter asked the questions in an individual random order. 10 questions involved spatial relations only including potential landmarks and 10 additionally addressed knowledge about the track configuration. The answering options were 'yes', 'don't know' and 'no'. 10 questions were answered correctly with a 'no' and 10 questions with a 'yes'. A correct 'yes' and correct 'no' were evaluated as correct answers. A wrong 'yes', a wrong 'no', and 'don't know' were evaluated as wrong answers.

3.2.2. Sketch-Map Task

After completing the relation questions, participants were asked to sketch the map on a sheet of paper. The frame defining the dimensions of the map was printed on the paper. The researcher and an independent rater evaluated all 48 sketches. The sketch maps were rated in two respects: (a) to what degree does the sketch resemble the original map concerning the course of roads, their parallelism and the junctions they have and (b) to what degree are potential landmarks represented in the correct position? The rating was performed on a 5-point Likert-type scale. A rating of 1 is associated with 'does not reflect the original' and 5 is associated with 'reflects the original precisely'. In the next section, we report a sum of the two ratings for better comparability with the results of the relation questions. See Figure 3 for an example sketch map.

3.3. Aggregate-Level Comparison: Convergent Validity

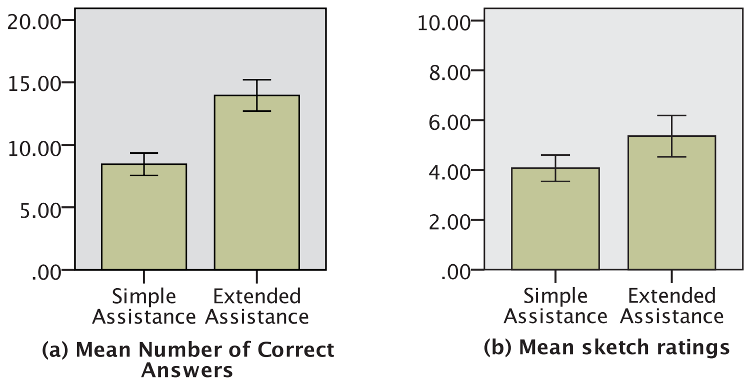

Figure 2. The mean results of (a) the relation-questions task and (b) the sketch ratings depending on the assistance condition under that the map was learned. The error bars show the 95 % confidence interval of the mean.

Regarding the number of correct answers to the relation questions, participants performed significantly better after having learned the map under the extended-assistance condition ($t(23) = 8.08$, $p < .001$, $r = .86$) and the effect is large. The combined ratings for the landmark and the track scale of the sketches show that participants' sketch maps were better given extended assistance ($t(23) = 2.79$, $p = .01$, $r = .50$), and the effect is fairly large. See Figure 2 for a visualization of the results.

The two tests showed comparable results for the main finding: Participants performed better when they learned the map under the extended-assistance condition. However, the relation-question task showed a comparably larger effect.

3.4. Comparison of Data-Point Pairs

As described, for each learning condition, each participant answered to questions concerning the spatial layout and was asked to produce a sketch map. For the purpose of comparison, the results of these two tests were evaluated in the following way: Each of the sketches was analyzed by the researcher. If the sketch clearly reflected spatial knowledge that was also asked for in one of the questions, the answer to that question reflected in the sketch was noted in a table. After this was done for all sketches, it was counted whether the answer to the questions based on the spatial layout of the sketch agreed with the one the participant gave to the verbal questions.

An example can illustrate the evaluation method. Question 19 (see Table 1) addresses whether a street called 'Amselweg' intersects with a street called 'Blumenstraße'. In the sketch map, these two streets do not intersect. See Figure 3 for an illustration, the relation is indicated by the arrow labeled with the number 19. As the figure shows, the participant did not draw the streets intersecting. A visualization of the original map is shown in Figure 4. As can be seen, in the original map the streets do not intersect.

If the sketch map and the relation questions both reflect the spatial mental model in an equal fashion, it is reasonable that the same participant should also have answered the relation question 19 with 'no'. If the participant did, we counted this pair of data points as equal. Pairs of data points with questions that were answered with 'don't know' were omitted from the evaluation.

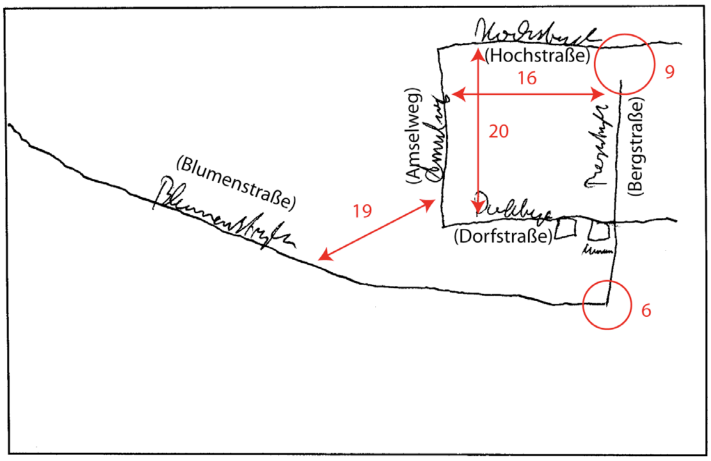

Figure 3. Scan of a sketch map produced by a participant, augmented by additional information. Data points that reflect knowledge that was also asked for in the relation questions are indicated by arrows and circles with the number corresponding to the question (see Table 1). A visualization of the original tactile map is shown in Figure 4.

Table 1. Translations of some of the questions posed in the experiment. The numbers correspond to knowledge reflected at the highlighted and numbered parts in Figure 3.

Question Number	Translation of the Question
6	Do Blumenstraße and Bergstraße form a t-crossing?
9	Do Hochstraße and Bergstraße intersect?
16	Do Bergstraße and Amselweg form a t-crossing?
19	Do Amselweg and Blumenstraße intersect?
20	Do Hochstraße and Dorfstraße meet?

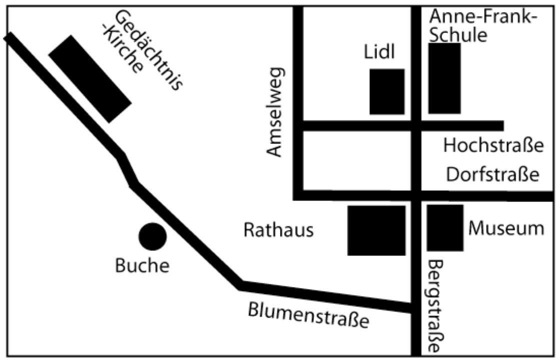

Figure 4. The original map for the sketch shown in Figure 3.

3.5. Results

Applying the procedure described above resulted in 504 pairs of data points that are compared. The data points agree to 87.70 %, which is a fairly high agreement (ϕ = .51, ϕ_{max} = .97, ϕ / ϕ_{max} = .53[3]). This indicates that the variation among methods is basically a result of how participants deal with information they are not sure about, i.e., a question about to what degree and how the task forces them to make a decision.

[3] Note that the ϕ / ϕ_{max} coefficient is not robust in the present case (extreme marginal values in the contingency table) [cf. 20].

4. Discussion

We compared the results of two outcome variables of an experiment that was conducted to investigate the effect of different assistance conditions. This comparison was performed on an aggregate level comparing averaged results and on the level of comparison of pairs of individual data points.

Regarding the aggregate level, ratings of the quality of the sketch maps and the verbally posed relation questions provide comparable results, even though the relation questions indicated a larger improvement of performance when the map was learned under the extended assistance condition than the sketch ratings did. When individual paired data points are compared to each other, the results indicate that both methods are valid measures of the internal spatial mental model of humans. An agreement of 87.70 % of the paired data points can be regarded as high agreement, taking into account that participants might have guessed about the spatial layout in both tasks when they were not sure about what to answer or how to sketch a part. This result supports the interpretation that both sketch maps and verbal assessment methods are valid assessment methods to be used in measuring the accuracy of the spatial mental model, for example, to compare the performance of different maps or geographic information systems.

However, it can only be speculated about the source of the remaining variation between the two methods. Even though it is likely to be due to guessing, other explanations that take the demands of the tasks into account are possible. The reported result has to be seen in the context of the fact that in the relation-questions task very coarse-grained, qualitative knowledge was needed to answer the questions posed. Therefore, the comparison reported here is an evaluation of the degree to which the two methods can be used to measure coarse-grained spatial survey knowledge. Used in this context, the effects of individual sketching ability and the different demands of the tasks do not affect the validity of the methods to a high degree.

References

[1] E. C. Tolman, "Cognitive maps in rats and men," *Psychological Review*, vol. 55, no. 4, p. 189–208, 1948.

[2] B. Tversky, "Cognitive maps, cognitive collages, and spatial mental models," in *Spatial Information Theory: A Theoretical Basis for GIS*, A. U. Frank and I. Campari, Eds. Berlin: Springer, 1993, p. 14–24.

[3] M. Hegarty, D. R. Montello, A. E. Richardson, T. Ishikawa, and K. Lovelace, "Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning," *Intelligence*, vol. 34, no. 2, p. 151–176, 2006.

[4] K. Lohmann, C. Eschenbach, and C. Habel, "Linking Spatial Haptic Perception to Linguistic Representations: Assisting Utterances for Tactile-Map Explorations," accepted for presentation at the Conference on Spatial Information Theory: COSIT '11, 2011, Belfast, MA.

[5] K. Lohmann, M. Kerzel, and C. Habel, "Generating Verbal Assistance for Tactile-Map Explorations," in *Proceedings of the 3rd Workshop on Multimodal Output Generation 2010*, Dublin, 2010.

[6] C. Habel, M. Kerzel, and K. Lohmann, "Verbal Assistance in Tactile-Map Explorations: A Case for Visual Representations and Reasoning," in *Proceedings of the AAAI Workshop on Visual Representations and Reasoning 2010*, 2010.

[7] R. D. Jacobson, “Cognitive mapping without sight: Four preliminary studies of spatial learning,” *Journal of Environmental Psychology*, vol. 18, p. 289–306, 1998.
[8] N. Newcombe, “Methods for the study of spatial cognition,” in *The Development of Spatial Cognition*, R. Cohen, Ed. Hillsdale, NJ and London: Lawrence Erlbaum, 1985, pp. 277-300.
[9] M. May, *Mentale Modelle von Städten*. Münster: Waxmann, 1992.
[10] K. Lynch, *The Image of the City*. Cambridge, MA; London: MIT Press, 1960.
[11] S. M. Kosslyn, K. H. Heldmeyer, and E. P. Locklear, “Children’s drawings as data about internal representations,” *Journal of Experimental Child Psychology*, vol. 23, no. 2, pp. 191-211, 1977.
[12] P. Jansen-Osmann, “Kognition von Distanzen - laborexperimentelle Untersuchungen in virtuellen Umgebungen,” Dissertation, Gerhard-Mercator-Universität Duisburg, 1998.
[13] R. Kitchin and M. Blades, *The Cognition of Geographic Space*, London; New York, I.B. Tauris, 2002.
[14] D. Wood and R. Beck, “Janine Eber maps London: Individual dimensions of cognitive imagery,” *Journal of Environmental Psychology*, vol. 9, no. 1, pp. 1-26, 1989.
[15] D. R. Montello, “The measurement of cognitive distance: Methods and construct validity,” *Journal of Environmental Psychology*, *11*(2), 101-122, 1991.
[16] M. Blades, “The reliability of data collected from sketch maps,” *Journal of Environmental Psychology*, vol. 10, no. 4, pp. 327-339, 1990.
[17] M. Billinghurst and S. Weghorst, “The use of sketch maps to measure cognitive maps of virtual environments,” in *Proceedings of the Virtual Reality Annual International Symposium*, 1995, pp. 40-47.
[18] A. W. Siegel and S. H. White, “The development of spatial representations of large-scale environments,” *Advances in Child Development and Behavior*, vol. 10, pp. 9-55, 1975.
[19] W. Wen, T. Ishikawa, and T. Sato, “Working memory in spatial knowledge acquisition: Differences in encoding processes and sense of direction,” *Applied Cognitive Psychology*, 2010.
[20] E. C. Davenport and N. A. El-Sanhurry, “Phi/Phimax: Review and Synthesis,” *Educational and Psychological Measurement*, vol. 51, no. 4, pp. 821-828, 1991.