

What Good Is Polarizing Research Into Qualitative and Quantitative?

by Kadriye Ercikan and Wolff-Michael Roth

In education research, a polar distinction is frequently made to describe and produce different kinds of research: *quantitative* versus *qualitative*. In this article, the authors argue against that polarization and the associated polarization of the “subjective” and the “objective,” and they question the attribution of generalizability to only one of the poles. The purpose of the article is twofold: (a) to demonstrate that this polarization is not meaningful or productive for education research, and (b) to propose an integrated approach to education research inquiry. The authors sketch how such integration might occur by adopting a continuum instead of a dichotomy of generalizability. They then consider how that continuum might be related to the types of research questions asked, and they argue that the questions asked should determine the modes of inquiry that are used to answer them.

A polar distinction is commonly used to describe and produce different kinds of research: *quantitative* versus *qualitative*. On the basis of our experience of teaching research design and methods and working with researchers, we believe the polarization is confusing to many and tends to limit research inquiry, often resulting in incomplete answers to research questions and potentially inappropriate inferences based on findings. This familiar polarization in education research is applied to types of research design, to data collection and construction, and to the methods of analysis used.

In this article, we argue against the polarization of research activities into qualitative and quantitative and the associated polarization of the “subjective” versus the “objective” and the attribution of generalizability to only one end of the polarity. We aim to demonstrate that this polarization is not meaningful or productive for education research. Our demonstration involves discussion of the following:

- *The problem.* We describe and discuss polarization as a fundamental problem in education research.
- *Existence of qualitative and quantitative characteristics in phenomena.* We demonstrate that the material world (ontology) and knowledge about it (epistemology) have both qualitative and quantitative characteristics.
- *Objectivity and subjectivity in constructing data.* We argue that the properties of “objectivity” and “subjectivity” associated with quantitative and qualitative research, respectively,

are neither accurate nor useful distinctions. We show that decision-making processes in qualitative and quantitative research follow the same interpretation model.

- *Generalizability.* The notion of generalizability is often, but incorrectly, associated with knowledge constructed through quantitative methods, and its lack with qualitative methods.

The problematic polarization of research into qualitative and quantitative has implications for the expected value and utility of various types of research and the generalizability of inferences that can be made on the basis of those types of research. In the penultimate section, we suggest moving beyond the quantitative–qualitative and associated dichotomies and thereby prepare the ground for more fruitful collaborations of researchers. We propose an integrated approach to education research inquiry. First, we recommend a conception of a continuum instead of a dichotomy. We show how this continuum accommodates notions of generalizability and relates to types of research questions asked and how these questions should determine modes of inquiry. Second, we emphasize focusing on research questions. And third, we suggest how investigators may accomplish the integration of different modes of inquiry by collaborating with researchers with expertise in those different modes.

The Problem

The quantitative–qualitative dichotomy not only distorts the conception of education research but also is fallacious. Philosophers of quite different ilk (e.g., Hegel, 1977; Husserl, 1968) agree that the material world has both quantitative (continuous) aspects—for example, temperature variation within a liquid—and discontinuous (qualitative) aspects—for example, boundary between liquid and solid phases of a substance. As discussed further in this article, there are quantitative and qualitative notions that describe and explain nature. The quantitative aspect of so-called qualitative research is not limited to summarizing findings but is inherent in all aspects of research categorized under the “qualitative” label. For example, in classroom observation, there are instances of student–student and student–teacher interaction that can be noted not just for type of interaction but for frequency as well. Similarly, in “quantitative research,” judgments about qualities and categorizations are needed not just when we interpret findings but in the data creation and collection stages as well. Variables used in statistical analyses, such as student self-esteem or mathematical competency, have both categorical–qualitative features (definition of variable) and quantitative–numerical features (continuous nature of representations).

The polar categorization of research in terms of the quantitative–qualitative distinction contributes to promoting research that emphasizes a certain type of data collection and certain construc-

Educational Researcher, Vol. 35, No. 5, pp. 14–23

tion modes rather than focusing on the construction of good research questions and conducting of good research. Quantitative researchers who may use statistical techniques for identifying differences, such as gender differences in mathematics, may leave it up to other researchers who will do qualitative studies to answer the question why. These different research agendas have been determined by the data construction methods that were adopted by the researchers. Similarly, researchers committed to using categorical data may not fully interpret their data and therefore may pursue research without the benefit of statistically summarizing and describing their data (Shaffer & Serlin, 2004).

A central concept (and practice) distinguishing education research is that of *generalization* and the associated notion of *generalizability*. Traditionally, quantitative research has as its goal to make claims about an entire population of cases on the basis of a subset of the population. The population most often consists of people ("Grade 8 students," "African Americans") but may also consist of events (Shaffer & Serlin, 2004). One can characterize this class of research, therefore, as making inferences of a certain type, from a sample of specimens to the entire population from which the specimens derive. The statistical apparatus provides a set of procedures to assess such issues as the probability of making an erroneous inference about the population. Thus, if a statistical comparison in a study suggests a difference between samples of Francophone and Anglophone students in Canada on some test and the statistical result is associated with a value of $p < 0.01$, the researcher knows that there is a probability of less than 1% that the difference between samples is mere coincidence. (We assume that all the assumptions of the research design and statistical method are met.) In other words, there is a probability lower than 1% of being wrong in saying that Francophone and Anglophone students performed differently for whatever reason.

Much of the research classified as qualitative differs with respect to this point of generalization. The approach of qualitative studies is to produce *thick descriptions* and, depending on the researcher, to generate a hierarchy of categories that summarize these descriptions. Because such hierarchies of categories are more economical than the descriptions, they constitute a form of theory that describes only the situation observed. The theory initially explains only the situation within which it is *grounded*. Some researchers *only* generate thick descriptions; others are more interested in the grounded theory. In this type of research, the level of inference is low. However, some level of interpretive inference is still required in order to bridge the gap between the sample of lived situation and the things that can be said about it.

But not all research that would fall into the currently used qualitative category restricts itself to describing a sample and inferring sets of patterns—that is, generating a grounded theory. For example, in the phenomenological tradition, some scholars are not simply interested in describing the experiences of one or more (small group of) people but, rather, attempt to make an inference about the more general conditions that make *different* experiences possible. Thus when Edmund Husserl (1980) investigated the experience of time, he was not interested in describing a (singular) experience of time but in describing the conditions that bring about time-related experiences in general. Critical psychologists (e.g., Holzkamp, 1983) also practice a type of qualitative research in which generalization is handled very differently from that in

other research falling under the same label. Grounded in Russian psychology of the Vygotsky–Leont'ev lineage, critical psychologists conceive of cultural–historical phenomena (including knowing and learning) in terms of the unity of quantity and quality, which leads to the notion of the *concrete universal*. Thus each observation (case) is understood to constitute a concrete realization of a possibility that exists at the collective level (population). Each observation therefore is simultaneously particular and universal, concrete and abstract, or specific and general.

On the other hand, in most quantitative education research, inferences do not generalize beyond the sample or contexts used because it is not feasible to conduct experimental designs or use random samples of students or schools. For example, it is not possible to draw students randomly to include in intervention programs, and researchers often need to resort to volunteers. Most correlational research does not use random, representative samples, and the results cannot be generalized beyond the range of data used in the analyses. In fact, many instances of research that use numerical data and a statistical approach for analysis are for descriptive purposes only. Generalizing the findings is severely limited when comparing groups that are not randomly equivalent and analyzing associations that are not based on data that cover the full range of possible variability. Shaffer and Serlin (2004) propose intra-sample statistical analysis as a technique for using statistical analysis to summarize data in qualitative research. Use of numbers and statistical methods is not limited to making high-inference interpretations such as generalizing beyond the sample or causal inferences. Rather, there are elements of quantification and generalization within qualitative research designs.

Within psychology, questions are raised about the generalizability of findings based on the typical experimental design. Researchers question whether generalizability actually is attained in most quantitative psychological research based on interindividual variation (Molenaar, 2004). Thus, "only under very strict conditions—which are hardly obtained in real psychological processes—can a generalization be made from a structure of interindividual variation to the analogous structure of intraindividual variation" (p. 201).

In answer to the problems raised, we elaborate in the sections below on the unity of quality and quantity in phenomena in different disciplines. We describe the interpretation models and the judgmental decision-making processes in constructing data to represent typological or topological aspects of phenomena. We propose a different concept of generalizability, which is necessary to demonstrate that some types of research classified as qualitative lead to generalizable inferences. The typical polarizations of research activities into qualitative and quantitative based on assumptions about the potential for generalization do not, in fact, serve worthwhile purposes.

Existence of Qualitative and Quantitative Characteristics in Phenomena

The classification of research activities into qualitative and quantitative typically is based on the types of data and analyses used. The analytic methods are ways of summarizing and describing data and making inferences systematically. Therefore, the main determining factor of this classification is the nature of the data. Data are *representations* of phenomena in nature, society, education,

and culture. Research activities are polarized into qualitative and quantitative classifications based on how phenomena are represented. In this section we show that natural and cultural phenomena in general (including the cognition of researchers and their participants) are simultaneously quantitative and qualitative. Full investigations of phenomena need to consider both of these aspects; therefore, it makes little sense to set up a qualitative–quantitative dichotomy in research.

In education research, as in the dominant mode of Western thought more generally, quantity and quality are treated as two independent, dichotomous phenomena, as different kinds of things (thereby pertaining to ontology) and different forms of knowledge (thereby pertaining to epistemology). The adjectives “quantitative” and “qualitative” are used in education, for example, to portray two distinct and apparently incompatible approaches to doing research (e.g., Lincoln & Guba, 1985). But not all scholars view quantity and quality as dichotomous entities. There are qualitative and quantitative differences in the world, even though, as radical constructivists rightly emphasize, we can never know what they really are. Nonetheless, they require qualitative and quantitative equivalents in knowledge and understanding (e.g., Thom, 1981).

In Perception and Its Mathematical Models

Fundamental to all research—whether quantitative or qualitative—is perception. Perceptual processes, however, are inherently quantitative and qualitative at the same time. Everything a qualitative researcher perceives also has a quantitative aspect; everything a quantitative researcher perceives also has a qualitative aspect. Thus, although light rays falling onto the retina are continuous and although the retina itself is a two-dimensional expanse, we see surfaces surrounded by edges (qualitative difference). The fact that in some cases we see edges, where neighboring points on the retina are dissociated, while in other cases we see continuous surfaces, where neighboring points fuse (Husserl, 1968), is a typical case of the unity of quantity and quality. Similarly, the reason that we hear as “bear” or “bare” the single sound stream transcribed as “b□ə” (according to the conventions of the International Phonetics Association) can be modeled using forms of mathematics that integrate quantity and quality (e.g., catastrophe theory [Thom, 1981]). Both the material stimuli and the interface between inside and outside worlds (e.g., the retina) are continuous; yet human beings perceive qualitative similarities and continuous (color) changes within surfaces and qualitative differences at their edges. In fact, qualitative discontinuities are salient in perception only if they are contiguously unfolded against the background of a moment that varies continuously and only if they present a sufficient gap or threshold of discrimination (Petitot, 1994).

In Philosophy, Political Economy, and Psychology

In dialectical philosophy, there is a tradition of considering quantity and quality as two negating moments of one and the same indivisible unit, as exemplified in the notion and practice of *measure* (Hegel, 1969). Thus any *measure*, any graph used in statistics, simultaneously involves quality (the nature of the variable) and quantity (its variation). Equally, an interpretive researcher who writes about a person being *more* or *less* competent, or *more* or *less* sorrowful, uses both quality (a type of competence, emotion) and

quantity (more, less). However determined, an assessment about the degree of something pertains to the quantitative domain. The identity of quality and quantity was a central feature of Marx’s (1976) political economy, where it figures both in the material world and in human consciousness. Beginning with the analysis of a commodity, which simultaneously has a qualitative moment (use value—exchanging grain for cloth) and a quantitative moment (exchange value—exchanging 2 bags of grain for 5 feet of cloth), Marx provides a historical analysis of how capitalist markets evolved in a continuous process from initial exchanges of goods that existed earlier in history. Central to his method was the continuous transformation of quality into quantity and quantity into quality. Using a similar, *historical genetic* method, critical psychologists showed how motivation and emotion may have emerged and developed, first in evolutionary and subsequently in cultural–historical processes that continuously transform quantity into quality and vice versa (e.g., Holzkamp, 1983).

In the Natural Sciences

In the natural sciences, there are many highly illustrative examples of the transitions of quantity into quality and vice versa. The qualitative differences during phase transitions (e.g., ice [solid] to water [liquid], water to vapor [gas], or ice to vapor) can be represented as a function of two variables, for example, temperature and pressure (Figure 1).¹ Take the substance at a particular pressure, corresponding to that at the level of the dotted line. If we keep one continuous variable constant—pressure—while heating the substance to increase its temperature continuously, then at some point (A, Figure 1) a small *quantitative* change in temperature brings about a *qualitative* change from solid to liquid. The best-known examples of the unity of quantity and quality in physics come from quantum mechanics, where the constituents of nature are represented by continuous differential equations but where observations are discontinuous and qualitatively different. For example, the functioning of a light meter in an SLR camera can be understood only by thinking of light in terms of (discrete) particles, but the refraction of light in the camera lens requires thinking of it in terms of (continuous) waves.

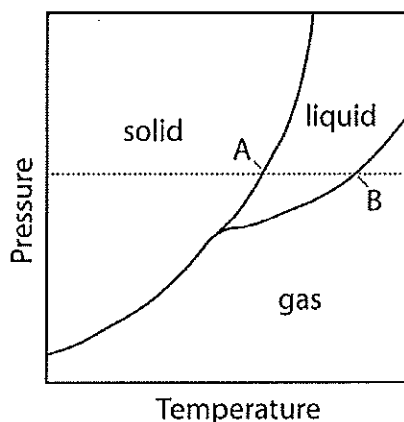


FIGURE 1. This generalized phase diagram of a substance shows how continuous (quantitative) variation of one variable, while other variables are held constant, leads at certain points (A, B) to qualitative changes in the substance (e.g., solid to liquid, liquid to gas).

In Classification and Judgment

Classification centrally involves perceptual phenomena that possess, as we have shown, both qualitative and quantitative characteristics. To understand the processes involved in student responses to test items, we therefore need to understand how the same material stimuli (ink dots on paper) are mapped onto qualitatively different perceptions and meanings. This is especially important on cross-cultural and cross-language tests. For example, when Francophone and Anglophone Canadian students read an item on an international test, they essentially are involved in a classification problem, where they segment a continuous perceptual stream into qualitative differences. Whether the French and English texts lead to similar segmentations cannot be taken for granted a priori. Also, when in our research four bilingual experts studying English and French versions of a test attempt to come to conclusions whether the two are qualitatively (sense) and quantitatively (degree of meaning) the same, they are involved in categorization, which, as shown above, is a unitary process in which quality and quantity are mutually constitutive. Quantitative research relies on data captured through surveys, tests, and checklists. These tools capture representations of a phenomenon at a discrete point in time. Yet the phenomenon itself is not discrete. For example, opinions captured in a survey at a fixed point in time are expected to vary across time and contexts. Typically, we are interested in measuring a latent construct that is continuous but may be probabilistically related to how one responds to the survey questions. Item response theory (IRT) describes relationships between continuous latent constructs and their discrete representations in the form of observations, responses, and answers (Steinberg & Thissen, 1996). IRT models are mathematical representations of the relationships between continuous latent constructs with discrete scores on a test, category of responses to survey questions, and categorization of observations.

Objectivity and Subjectivity in Constructing Data

Associated with the polarization of research into quantitative and qualitative is another, correlated polarization, according to which the former type of research is objective and the latter subjective. Although qualitative research is thought to be context based and the inclusion of the researcher's subjective perspective enriches the quality of the research, quantitative research is considered to be objective and its judgments are expected to be replicable by other researchers. Even though these approaches to viewing the role of the researcher and the research activity seem very different, both types of research activities involve subjective judgments. In quantitative research, once the data are constructed, statistical methods constrain and define the types of inferences that can be made on the basis of the data. Yet there are many stages of the data construction that require "subjective," "defensible" judgments by the researcher (Ercikan & Roth, 2006).

Most common forms of data that are used in quantitative research use tests, measures, or surveys. We can understand these tests as tools by means of which researchers attempt to capture participants' knowledge, opinions, feelings, and so on. All of these are constructs that are not directly observable. The tools provide the means of gathering material evidence (sometimes referred to as "raw data") used for the construction of data (the entities used in support of research claims, findings) with regard to the degree of or the nature of the constructs. Measurement tools include questions or prompts

that require participants to report, create, or identify responses. The participant responses in and of themselves do not have any meaning in relation to the construct. For example, participants' description of a solution strategy to a mathematics test question cannot be considered as evidence of mathematical competency without considerations of what is meant by mathematical competency. That is, whether the proposed solution strategy is one of the correct strategies, whether it uses given information fully, and so forth.

In typical testing contexts, these considerations are taken into account through an interpretation model described by a set of scoring rules. These determine what is important to pay attention to in the student's response, what kind of evidence the response provides regarding the competency, and the significance of each aspect of the response in relation to the mathematical competency that one is interested in assessing (Mislevy, Wilson, Ercikan, & Chudowsky, 2002). If the mathematics competency definition includes problem solving, the scoring rule would assign credit for simply stating an accurate strategy for solving a mathematics problem, whereas if the problem solving is not part of the definition of mathematical competency, stating an accurate strategy may not earn the student any credit if the final calculation does not lead to a correct answer.

In research that uses quantitative data, many subjective judgments are made. The first judgmental aspect of constructing data in the example above is selection of an interpretation model. Researchers need to make many decisions regarding the interpretation model: What aspects of students' responses provide evidence of competency? What are the best ways of assigning scores to different aspects of responses? Integrated assessment tasks are good examples for elaborating how the interpretation model may be very different depending on the definition of the construct. In such tasks, students' responses are evaluated for evidence of multiple competencies. For example, the interpretation model for assigning the maximum score of 3 for writing evaluates several competencies in a student's writing: "development," "organization," "attention to audience," and "language." An interpretation model for science might evaluate responses for evidence of knowledge acquisition and integration of major concepts and unifying themes from earth or space sciences.

The second level of subjectivity is present in judgments when scorers apply the scoring rubrics to students' responses or products. The scorers are guided by the scoring rules but use their informed judgments to evaluate the evidence provided by responses. The more structured the scoring rubrics are, the more consistent the scorers' judgments will be. However, inconsistencies among scorers are inevitable because of the judgmental nature of the scoring process. One scorer may think that there is sufficient demonstration of a particular skill to grant the response a score of 3 while another scorer may decide that the response lacks the degree of detail needed to confidently assign the response a score of 3. Such differences tend to disappear when scorers have experiences of scoring with others and of discussing any differences (Schoenfeld, 1992). Therefore, the subjectivity involved in this phase of data construction depends on who is involved in the scoring, their level of experience in scoring, the scoring rules, and the student responses or products.

The two phases of data construction described above are associated with a particular test. There are many subjective decisions

and choices that a researcher who uses quantitative data makes prior to and during the development of the test. These include but are not limited to the definition of the competency, item types and formats, contexts the items will be presented in, and the actual items that will be created. All of these points of judgmental decision-making are sources of deviation for researchers who are interested in replicating the research study and the data construction processes. The judgments involved in these decision points are potential sources of inconsistency (error) and, therefore, may potentially limit generalizability of findings across research studies.

The interpretation models used in scoring student responses to test questions are similar to interpretation models used in research that involves observations, videotapes, interviews, and so forth. These are data sources that can reveal evidence of constructs, events, and interactions; but they require an interpretation by the researcher of what is happening in a student–student or student–teacher interaction in a classroom. Any data that are constructed arise through an interpretation model that involves subjective judgments. The processes involved in constructing data may be quantitative or qualitative in nature and include three dimensions: data sources, interpretation model, and data. *Data sources* can include response data to test questions, interview transcriptions or videotapes, or video- or audiotapes of classroom processes (Figure 2, left). The *interpretation model* can include scoring rules, coding protocols, or filters for extracting relevant data from the data sources (Figure 2, center). The resulting *data* (Figure 2, right) can be constructed by applying the interpretation model in the form of scores on test questions, types of expressions and events, or numbers of expressions and events. This model articulates how and why *all* forms of inquiry—irrespective of the adjectives “quantitative” and “qualitative”—involve subjective and objective, reproducible and verifiable, moments. Moving beyond a polarization of research therefore also requires superseding the quantitative–qualitative dichotomy with a new approach, which we articulate in terms of rethinking generalizability.

Generalizability

Thus far we have argued that the quantitative–qualitative dichotomy is not appropriate for distinguishing forms of education

research because (a) all phenomena are quantitative and qualitative at the same time; and (b) data construction processes follow similar interpretation processes for all education research; and (c) for most constructs that education researchers are interested in, these data construction processes are based on subjective, defensible judgments. There is, however, another dimension that is often used to distinguish types of education research. The discussion of degrees of *generalizability* is inherently associated with statistical (i.e., quantitative) approaches and commonly questions the generalizability of observational (i.e., qualitative) approaches. Some researchers, especially those knowledgeable about research methods associated with both approaches, will agree with us that generalizability is not limited to “quantitative research”; however, generalizability has a specific statistical definition and is commonly considered relevant in quantitative research only. In this section, we articulate why this association is problematic. We describe generalization in two areas of research, which employ methods that on the surface look qualitative but that strive for generalization beyond the particular sample studied.

Generalization in Experimental Phenomenology and Neurophenomenology

We exemplify different levels of abstraction and generalization in phenomenological and neurophenomenological approaches with a case from perception that readers can easily reproduce themselves. Figure 3 represents the type of images researched by the Gestalt psychologist Edgar Rubin. Most people perceive the image as a cross that is oriented along the diagonals rather than as a broad-leafed Maltese cross, which in fact constitutes the ground for the former (Spillmann, 1999). Gestalt theorists explained the phenomenon in terms of the *law of proximity*, according to which items that are closer are grouped preferentially. In the present situation, the cross oriented along the diagonals is perceived preferentially rather than the upright, broad-leafed Maltese cross.

With some practice, viewers may see the upright Maltese cross as figure. This is a qualitatively different perception, although the material stimulus is the same. Whichever cross we perceive, the area within the figure is continuous white. Yet other perceptual experiences are also possible—for example, a square cir-

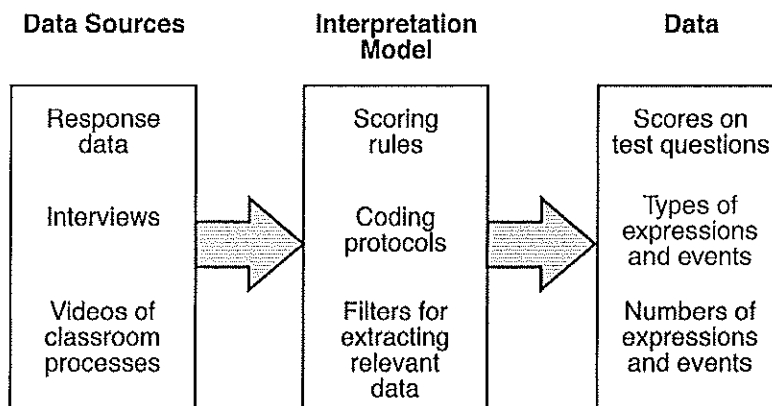


FIGURE 2. In the process of constructing data that enter into a statistical analysis, judgments—and therefore qualitative dimensions—enter into “quantitative” research.

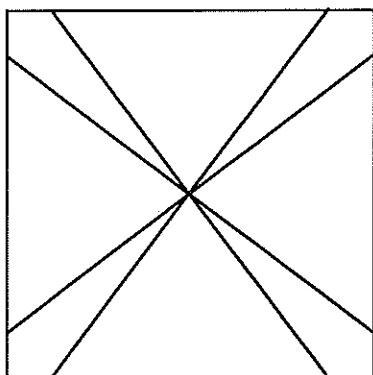


FIGURE 3. This Maltese cross configuration is used in gestalt psychological investigations and has contributed to the establishment of a law of proximity.

cus tent viewed from above. A researcher interested in describing all the different perceptual experiences that exist within a particular group of figures is conducting a *phenomenographic* study (cf. Marton, 1994). However, it is not necessary to interview many people; one individual can reproduce the full range of perceptual experiences (Ihde, 1986). Whether the evidence from an individual or a group of individuals was presented, this would be research of the low-inference type. However, phenomenological philosophers and natural scientists (Varela, 2001) are not satisfied stopping here but want to generate statements about experience *in general*. Answering a question such as “What is at the source of these different experiences?” seeks to generalize from the different experiences of *one* individual to the possibility of *these* human experiences *in general*. Thus, returning to Figure 3, further inquiry shows that the reversal between figure and ground is due to shifts in focal point. Moving the focal point back and forth between points inside the narrow diagonal cross to a point inside the broad-leaved standing Maltese cross we can actively switch figure and ground and therefore between perceiving the two crosses.

Such perceptual experiences become the raw data for inference, the second stage in the analysis, which investigates the conditions under which differences occur. For example, we may realize that qualitatively different figure-ground constellations are functions not only of the locus of the focal point but also of (continuous) movement. In this particular case the movement is that of the eye itself when the head and body are fixed. One can notice how the eye jumps (saccades) between figure and ground, staying longer on the figure. After intense reflection on the processes, the phenomenological researcher engages in times of non-attention to the phenomena. Through this non-attention, the emergence of insights is provoked (Depraz, Varela, & Vermersch, 2002).

This example shows how the (neuro-) phenomenological method begins with (perceptual) experience but then engages in systematic inquiries (variation of experience that leads to variation in sensations) and descriptions thereof so that these can be repeated and validated by other researchers. These generalizations, even if based on the experiences of only one person, pertain to the experiences of many if not all persons in a culture or humankind in most general terms.

Generalization in Dialectical Materialist Approaches

Wittgenstein (1958) formulated a key problem with generalizing from properties in his examples of family resemblance. For example, Smith-A and Smith-B have the properties a, b, and c in common; Smith-B shares with Smith-C the attributes b, c, and d; Smith-D and Smith-A share property a; and Smith-E and Smith-A have no attribute in common but their name. There is therefore no single attribute that *all* Smiths share, although any subsample of Smiths may have one or more common attributes. Yet all the Smiths might be related in a more fundamental way if they are all part of the same family, for example, if Smith-A has a parental relationship to all other Smiths. Smith-A therefore embodies *all* the possibilities that are concretely realized in the filial generation; but each offspring realizes these possibilities in very different ways.

The underlying parental relationship between the different Smiths, despite the apparent differences, calls for a (cultural) historical—that is, genetic—explanation (Il'enkov, 1977). Formally, we are dealing with a genetic explanation when a certain fact is not derived from antecedent conditions and laws (deduction) or observations and antecedents (induction) but rather is shown to be the endpoint of a longer development, the individual stages (phases) of which can be followed. In *causal genetic* explanations, the antecedents of a step coincide with the outcome of the previous step. In historical genetic explanations, the antecedents of one step do not coincide with the outcome of the previous step, requiring new pieces of information to be added.

Historical-genetic explanations are important to understand. For example, the emergence of linguistic terms illustrates the different ways in which meaning can be made within and across cultures. The historical-genetic relationship of terms is important to understanding the degree to which different language versions of a mathematics test (e.g., English/French versions of tests in Canada, or the PISA tests) measure the same construct. The relationship of linguistic terms across and between cultures can be illustrated with the following example. The southern French only use the term *abeille* (derived from the Latin *apicula*) for bee, while the people of northern France use a variety of terms. One of these, *apis*, has been preserved in the French words *apiculture* (beekeeping), *apiculteur* (apiculturist, beekeeper), and so on. The semantic fields in different languages differ, especially in cases where words had different origins but also when words had the same origins. Thus, in English, *bee* has in its semantic field *busy bee*, *a busy worker*, and *to put a bee on*, the last of which means to put an end to and also to ask for a loan. In French, *abeille* has not only *bee* as a sense but also *sweatshop owner* and a hairstyle named *arrowhead*. Semantic fields not only differ, they also change; and they change differently in different countries. In German, for example, the equivalent of bee is *Biene*, once slang for *woman* in roughly the same sense as the English *chick* that was popular in the 1960s and 1970s. (This use has all but disappeared, but *Biene* remains as a diminutive of the given name Sabine. Among older people the association with *chick* continues to exist, whereas the middle school students of today may not even know that signification.) In German the term *Biene* also has associations with assiduity (like busy bee), and *Bienenstich* is both a bee sting and a type of cake.

This example embodies the fundamental ideas in a materialist dialectical and historical logic, whereby there is a mutually constitutive relation between the general (universal, abstract) and the

specific (particular, concrete). The original root word *apicula* in Latin is the universal, but it is as concrete as the specific words now used in French (*abeille*), Portuguese (*abelha*), and Spanish (*abeja*); the concrete universal root word *apis* develops into the Italian *ape* and continues to live in the English terms *apiculture* and *apiculturist*. In cultural–historical psychology, researchers therefore speak of the *ascension* of the abstract to the concrete, that is, the concrete universal to the concrete specific.

This approach leads us to different ways of thinking about knowing and learning (e.g., Vygotsky, 1986). Culture in general and language in particular constitute generalized possibilities but are available only through the ways in which they are actually realized. Both possibilities and realizations are concrete, although the former are general and the latter specific. Therefore, some cultural possibilities are not realized at a given moment although they do exist and are recognized as possibilities by individual subjects. A student’s (or evaluator’s) specific interpretation of a word problem may vary from the general. It is not deficient in any sense, however, but rather an expression of the other possibilities that cultural tools such as language provide. Because a word or tool does not contain instructions for its use, students cannot make an a priori decision as to the extent to which their use of a word or interpretation is the one that represents the interpretation of most people. A child’s explanation of the universe in terms of a moving sun and a stationary earth is not a defective perception of the world, as some research claims, but a discursive possibility arising from the everyday uses of expressions such as “The sun sets,” “What a beautiful sunrise,” or “Stars move across the sky.”

Moving Beyond Dichotomies

In the previous sections we demonstrated that polarization of research into qualitative and quantitative is neither meaningful nor reflective of the realities of research. Similarly, neither the associated subjective–objective distinction nor the association of quantitative research and generalizability are correct or useful. We therefore (a) make a proposal *toward an integrative framework*; (b) suggest a change in drivers by *putting research questions first*; and (c) encourage investigators to *join expertise and work together*.

Toward an Integrative Framework

We suggest that rather than distinguishing research by means of a dichotomous quantitative–qualitative dimension to locate different forms of research on a continuous scale that goes from the (unreachable) lived experience of people on one end (e.g., what doing and learning mathematics feels like to girls and boys) to idealized patterns of human experience (e.g., a significant correlation between short-term storage space and mathematical ability), on the other. Whereas previously, quantitative and qualitative constituted (qualitatively) different categories of research, they are now on the same scale and therefore only different by degree. The formerly distinct forms of quantitative and qualitative research now are located on different parts of the same scale, which is characterized by low levels of inference on one end and high levels of inference on the other end (see Figure 4). The scale simultaneously is continuous, because achieved in research praxis, and discontinuous, because it involves translations from representation to representation (Latour, 1999). This requires anyone who wants to make polar distinctions to identify where boundaries are to be drawn and on what grounds there ought to be boundaries that attribute research to different categories.

Knowledge derived through lower-level inference processes (left-hand direction in table) is characterized by contingency, particularity, being affected by the context, and concretization. Knowledge derived through higher-level inferences is characterized by standardization, universality, distance, and abstraction (right-hand direction in table). The more contingent, particular, and concrete knowledge is, the more it involves inexpressible biographical experiences and ways in which human beings are affected by the dramas of everyday life. The more standardized, universal, distanced, and abstract knowledge is, the more it summarizes situations and relevant situated knowledge in terms of big pictures and general ideas.

Descriptive research, which may involve qualitative or quantitative data, will fall toward the left end of the scale in Figure 4; grounded theory or group comparisons based on nonrepresentative samples, which require a translation and summary from the initial material evidence, lie farther to the right. For instance, when researchers make inferences to populations on the basis of

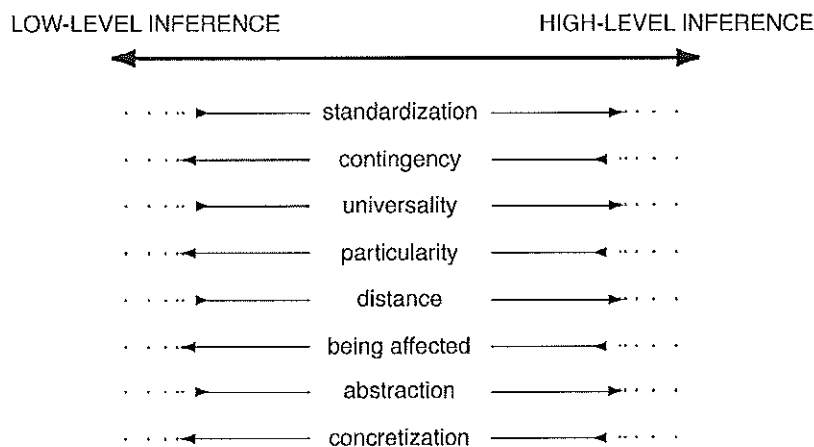


FIGURE 4. Continuum of low-level-inference to high-level-inference research and associated tendencies for knowledge characteristics along eight dimensions.

representative samples, those inferences lie closer to but not exactly at the right-hand end of the scale. The products (e.g., articles, representations) of both forms of research inherently re-present (aspects of) human experience. Both are abstractions, although of different scales, rather than constituting human experience itself. Irrespective of the type of research, both require an initial observational translation—that is, a reduction or abstraction that is prelogical and preconscious and leads to data.

The relational adjectives “high-level inference” and “low-level inference” allow us to place forms of research on the right that on their surface look qualitative even though they do not use mathematical representation, as our examples of phenomenological inquiry and dialectical materialist approaches showed. These research types inherently intend to generalize beyond the subjects and contexts they use. Justifications of such high levels of inference lie in the nature of the phenomena they study, for example, inferences about the commonality of human biological functioning or factors affecting visual perception.

Research at different ends of this continuum addresses different questions. In the simple examples given above, about mathematics learning of boys and girls, a question may be “Are there gender differences between boys and girls in mathematics learning?” A statistical analysis based on test data from a nationally representative sample of boys and girls may provide partial answers to this question. However, such an analysis would not consider how and what types of mathematics different groups learned or how the mathematics test that was used matched the learning contexts and content. An ethnographic case study may provide answers to questions about possible differences in socialization of girls and boys in mathematics classes through gendered variations in discursive practices. These two examples may very well place themselves at different ends of the scale on Figure 4.

Although there is value in such distinct types of research, we believe that research placing itself at either extreme end cannot be expected to fully answer research questions. Gender differences based on typical statistical analyses of group results are, in fact, differences in group means. These differences reflected in the group means may have very different patterns for different groups of students. Some students may have different mathematical skills, while others may have been taught different curricula using different instructional methods and their results do not provide any information about individual students. Therefore, the findings provide neither information about individual students nor insights about subgroups of students (e.g., high-performing students who may have differential patterns of gender differences) that are not the focus of statistical comparisons. The findings are relevant only for the overall group that the analyses focused on. To make educational decisions based on research findings, we need to look more closely at which groups of girls are performing differently from boys and what it is about being girls in an educational context that may lead to differential performance in mathematics (Walkerdine, 1988). Therefore, we need to know what aspects of mathematics education and what aspects of gender in a particular context (society) may be associated with mathematics learning and performance. Similarly, an ethnographic study of gender differences in the socialization of boys and girls in a particular mathematics class may be expected to provide limited guidance to educators. In this type of research, the researcher has the burden of demonstrating the relevance of the research for

other contexts. Research that examines representative groups for purposes of generalizability, as well as the individual and groups of individuals with certain characteristics, is expected to provide a more complete picture of the nature of and sources of differences and to provide better guidance for education decision makers and policymakers.

The meaningfulness and completeness of answers to research questions become the primary focus when we consider the purposes of education research that is intended to serve society. This meaningfulness and completeness are possible if researchers formulate the research questions first and let the research questions dictate and determine what modes of inquiry are more appropriate. The chosen methods may combine several approaches that are associated with different ends of the continuum. This requires researchers to have knowledge about a wide range of methodologies and possibly expertise in how different methodologies work together. In the subsections below, we discuss these two issues, which are critical to conducting research that produces meaningful findings by (a) putting research questions first, and (b) joining expertise as researchers work together.

Putting Research Questions First

The purpose of research is to generate knowledge rather than to concretely realize one method or another. Research methods are means to answer knowledge-constitutive questions. We therefore suggest that research questions, not method, ought to drive education research. This call for research questions as drivers of research is consistent with a report generated by a National Research Council committee charged with addressing three related questions: (a) What are the principles of scientific quality in education research? (b) How can a federal research agency promote and protect scientific quality in the education research it supports? (c) How can research-based knowledge in education accumulate? (Shavelson & Towne, 2002, p. 26). The report and various publications arising from it emphasize that research questions should drive the choice of research methods. The types of research questions asked therefore depend on the stage of development in the area of exploration.

Research can be classified according to three types of questions that it may answer: “What is happening?” “Is there a systematic effect?” and “Why or how is it happening?” These three questions, considered in sequence, locate research increasingly to the right in Figure 4. To address the first type of question we may use ethnography, phenomenography, case study, description of a population using a statistical sample, or interviews. The second type of research question explores causal relationships and requires some form of experimental or quasi-experimental design and causal model. We suggest testing the generalizability of such causal relationships through case studies. The third type of research question—“Why or how is it happening?”—explores a causal agent or mechanism. An iterative tryout-redesign-tryout approach may address replicability and generalizability, and advance claims for understanding the mechanisms.

The three types of research questions correspond to different levels of inference, from low to high. For example, in the first stage one could ask questions such as “How do girls and boys compare with each other in mathematics learning?” or “What are Canadians’ perceptions and understandings of their personal and national histories?” Both of these types of questions can be explored

by using methods, data construction approaches, and analyses that are associated with qualitative and quantitative research. Gender differences in mathematics are widely addressed by means of large-scale assessments of mathematics learning but could also be explored through ethnographic studies within classrooms or structured teacher interviews. This type of research will be descriptive in nature. Research that explores causal relationships or mechanisms intends to make higher levels of inference. Answering the question "How and why do girls tend to perform more poorly on large-scale mathematics assessments?" requires a research design that generalizes beyond what may be happening with individual learners or specific classrooms. Such a design targets high inference levels, yet probably cannot be addressed by carefully designed experiments alone.

Joining Expertise as Researchers Work Together

Current courses in education research methods tend to prepare researchers for conducting research at the ends of the continuum. Therefore, researchers with the best of intentions find themselves in situations where, even though they realize the limitations of their existing research program, they themselves are not prepared to extend the research that may require modes of inquiry different from those for which they are trained. For meaningful explorations and integration of different modes of inquiry, research questions and what the researchers would like to generate should dictate what research methods and modes of inquiry are used. In practice, this approach may require well-integrated collaborations among researchers with expertise in different modes of inquiry. To create meaningful findings, the collaborators should start working together from the beginning and during the entire research process.

In the early 1990s, two teachers set out to conduct an ethnographic study of knowing and learning in two eighth-grade science classrooms organized around the then-salient metaphors of cognitive apprenticeship, community of practice, and authentic science (e.g., Roth & Bowen, 1995). Each science class was videotaped; all artifacts that students produced were collected, and students were interviewed about salient aspects of the learning environment. The teacher-researchers also collected numerical responses to various learning environment questionnaires and student grades, which were used to construct correlational information about *this* student sample. The transcription of videotaped lessons and interviews was completed within 48 hours. During the transcription, early hypotheses were formed about how students know and learn. One of the hypotheses was: "Student groups who interpret the relationship between many data pairs are more likely to use mathematical-statistical approaches than students who worked with a small number of data pairs." The researchers focused on student groups (usually pairs), because this was thought to constitute a more ecologically valid method of research, as the students were working collaboratively for the entire curriculum unit. The researchers then designed a controlled experiment wherein three forms of a test, containing 5, 8, and 17 data pairs, respectively, were randomly distributed to students in two experimental classes and one control class at the same school. While students worked on the test, the researchers videotaped individual groups. The statistics showed that the hypothesis had to be rejected. However, the videotapes allowed the researchers to construct new knowledge and hypotheses

about the differences between students' science-and mathematics-related practices in contexts where they formulated the objects and goals of activity, as opposed to contexts where the teacher-researchers provided the objects and goals (Roth, 1996).

In this situation, the two teachers integrated two forms of research, often opposed and located at polar opposites. The advantage of this way of working is that the two teachers conducting the research are able to "translate" their research findings into classroom practice, because the low-level inference observations and categorizations were sufficiently close to the events that implications could be derived. At the same time, the high-level inferences are important for understanding ways of thinking more generally, not only across groups but also in other classrooms. The fact that the teachers used a controlled experiment allowed them to state generalizations that were likely to hold beyond their school. Together, the two researchers transcended the distinction between quantitative and qualitative research and generated knowledge simultaneously useful to classroom teachers and educational theoreticians. We expect such collaborations—combining differing degrees of *general* (context independent) and *specific* (contextual)—to lead to research that focuses on meaning, implementation, and inferences rather than on inappropriate adoption of methodologies married to opposite ends of the continuum of low-to-high-inference research.

Discussion and Conclusions

In answer to our rhetorical question "What good is polarizing research into qualitative and quantitative?" we provided arguments against polarizing education research. We began our arguments by identifying polarization as a problem and proceeded to articulate why it is problematic in three ways: (a) All phenomena and all knowledge simultaneously have quantitative and qualitative dimensions; (b) the distinction between objectivity and subjectivity, normally associated with that between quantitative and qualitative research, is neither accurate nor useful; and (c) generalizability is not a feature of mathematization but a descriptor for the tendency of inferences to go beyond the context and participants involved in the research. We then proposed moving beyond polarization by (a) using a different classificatory continuum based on the relational terms "low inference" and "high inference"; (b) emphasizing a focus on the research question; and (c) encouraging the collaboration of researchers with expertise in forms of research formerly labeled quantitative and qualitative.

In this call to transcend dichotomies, we simultaneously suggest reaching new forms of research that go beyond forms that can be placed on a single location on the scale. Rather, the aim of education research should be to produce research results that are characterized simultaneously by high and low inference levels, with results that include the standardizing, universalizing, distancing, and abstracting aspects of knowledge, as well as the contingent, particular, and concrete aspects. This integration would also allow us to transcend the distinctions that previously have been attributed to different knowledge-constitutive technical, practical, and emancipatory *interests*, respectively associated with (a) prediction, instrumental knowledge, and control; (b) interpretation and understanding; and (c) criticism, liberation, and reflection (Habermas, 1971).

The category of knowledge-constitutive interests has frequently been used to characterize experimental research in a negative way,

and qualitative and critical inquiry in a positive way. All research, however, involves abstraction, although at different levels of scale. We suggest focusing instead on the needs of different people in society, who require different forms of knowledge to make decisions. The politicians who need to make spending decisions about educational funding require different forms of knowledge than the teacher who has a child with spina bifida in her classroom. The decisions that these different (groups of) individuals make require knowledge that summarizes experience in different ways. Inherently, this places their knowledge needs at different levels on our scale.

Instead of dichotomizing research into qualitative and quantitative, we need integrative approaches that provide the appropriate forms of knowledge needed by decision makers located differently in society and dealing with different units of analysis (individual, group, community, etc.). In this approach, we need researchers to make choices regarding data sources, data construction, and analysis methods that best fit their research questions and to consider using multiple approaches and modes of inquiry.

NOTES

Research grants from the Social Sciences and Humanities Research Council of Canada enabled the writing of this article. We are grateful to Robert Anthony, who assisted us with his careful edits and advice. The author names at the head of the article are given in alphabetical order because the authors made equal contributions.

¹Only two variables are necessary, because they determine the system as a whole, as can be seen in the ideal gas equation $p \times V = \text{constant} \times T$. Given two variables, such as temperature (T) and pressure (p), this equation fixes the values of the third. No matter how the substance H_2O is represented, even radical constructivists have to acknowledge the qualitatively different experiences that we have when interacting with ice, water, and vapor—try walking on water!—and the continuity of human experiences with water as it changes temperature.

REFERENCES

- Depraz, N., Varela, F. & Vermersch, P. (2002). *On becoming aware: Steps to a phenomenological pragmatics*. Amsterdam: Benjamins.
- Ercikan, K., & Roth, W.-M. (2006). Constructing data. In C. Conrad & R. Serlin (Eds.), *SAGE Handbook for research in education: Engaging ideas and enriching inquiry* (pp. 451–475). Thousand Oaks, CA: Sage.
- Habermas, J. (1971). *Knowledge and human interests*. Boston: Beacon.
- Hegel, G. W. F. (1969). *Hegel's science of logic* (A. V. Miller, Trans.). London: George Allen and Unwin.
- Hegel, G. W. F. (1977). *Phenomenology of spirit* (A. V. Miller, Trans.). Oxford: Oxford University Press.
- Holzkamp, K. (1983). *Grundlegung der Psychologie* [Foundations of psychology]. Frankfurt, Germany: Campus Verlag.
- Husserl, E. (1968). *Logische Untersuchungen* [Logical investigations]. Tübingen, Germany: Niemeyer.
- Husserl, E. (1980). *Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins* [On the phenomenology of the consciousness of internal time]. Tübingen, Germany: Niemeyer.
- Ihde, D. (1986). *Experimental phenomenology*. Albany: State University of New York Press.
- IPenkov, E. (1977). *Dialectical logic: Essays in its history and theory* (H. Campbell Creighton, Trans.). Moscow: Progress.
- Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA: Harvard University Press.
- Lincoln, Y. S., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Marton, F. M. (1994). Phenomenography. In T. Husén & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed., Vol. 8; pp. 4424–4429). Oxford, UK: Pergamon.
- Marx, K. (1976). *Capital* (Vol. 1). London: Penguin Books.
- Mislevy, R., Wilson, M., Ercikan, K., & Chudowski, N. (2002). Psychometric principles in student evaluation. In D. Nevo & D. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 478–520). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, 4, 201–219.
- Petitot, J. (1994). Phenomenology of perception, qualitative physics and sheaf mereology. In R. Casati, B. Smith, & G. White (Eds.), *Philosophy and the cognitive sciences: Proceedings of the 16th International Wittgenstein Symposium* (pp. 387–408). Vienna: Verlag Hölder-Pichler-Tempsky.
- Roth, W.-M. (1996). Where is the context in contextual word problems? Mathematical practices and products in Grade 8 students' answers to story problems. *Cognition and Instruction*, 14, 487–527.
- Roth, W.-M., & Bowen, G. M. (1995). Knowing and interacting: A study of culture, practices, and resources in a Grade 8 open-inquiry science classroom guided by a cognitive apprenticeship metaphor. *Cognition and Instruction*, 13, 73–128.
- Schoenfeld, A. (1992). On paradigms and methods: What do you do when the ones you know don't do what you want them to? Issues in the analysis of data in the form of video-tapes. *Journal of the Learning Sciences*, 2, 179–214.
- Shaffer, D. W., & Serlin, R. C. (2004). What good are statistics that don't generalize? *Educational Researcher*, 33, 14–25.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Steinberg, L., & Thissen, D. (1996). Uses of item response theory and the testlet concept in the measurement of psychopathology. *Psychological Methods*, 1, 81–97.
- Thom, R. (1981). Worüber soll man sich wundern? [What should we wonder about?]. In K. Maurin, K. Michalski, & E. Rudolph (Eds.), *Offene Systeme II: Logik und Zeit* (pp. 41–107). Stuttgart: Klett-Cotta.
- Varela, F. J. (2001). Consciousness: The inside view. *Trends in Cognitive Sciences*, 5, 318–319.
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Walkerdine, V. (1988). *The mastery of reason*. London: Routledge.
- Wittgenstein, L. (1958). *Philosophical investigations* (3rd ed.). New York: Macmillan.

AUTHORS

KADRIYE ERCIKAN is an Associate Professor of Educational and Counselling Psychology and Special Education at the University of British Columbia, Faculty of Education, 2125 Main Mall, Vancouver, BC V6S 1Z4, Canada; kadriye.ercikan@ubc.ca. She specializes in research methods, cross-cultural and language issues in assessments, and psychometrics.

WOLFF-MICHAEL ROTH is Lansdowne Professor of Applied Cognitive Science at the University of Victoria, MacLaurin Building A548, Victoria, BC V8W 3N4, Canada; mroth@uvic.ca. He specializes in applied cognitive science, learning across the lifespan, cognition in mathematics and science, and cultural-historical theories.

Manuscript received April 6, 2005
Revisions received August 25, 2005, and
December 20, 2005
Accepted March 14, 2006