The problem of method: A philosophical analysis of the Standardized Test of Literacy

David Ian Hanauer
Indiana University of Pennsylvania, U.S.A.

The problem of method for the human sciences has been a source of contention and continued polemic within the framework of philosophical discussions of hermeneutics for many years (Gadamer, 2003; Grondin, 1994). Essentially the argument has focused on the possibility or impossibility of defining a systematic method of understanding for constructs such as human reading and writing. Within the current world trend for standardizing testing, directed by educational practices in the U.S., the problem of method for the human sciences has been relocated from the cognitive to the textual and thus essentially disregarded as irrelevant. While method is described as central to standardized testing by practitioners and employees of the various testing institutions, the term method is conceptualized within the constraints of a "testcentric" concept of testing that focuses on the test as a unitary, autonomous object (Hanauer, 2002). This testcentric concept of method is far removed from the actual problem of an interpretive method as discussed within philosophical hermeneutics. The current manuscript will attempt to address this relocation of the concept of method and by so doing retheorize the nature of the standardized test.

To date, the critique of standardized testing has mainly been conducted by the proponents of social criticism (Foucault, 1979; Madaus, 1990; McNamara, 1998; Shohamy, 2001a; Spolsky, 1997). In a series of analyses, both theoretical and empirical, the standardized test is seen to discriminate on a cultural and ethnic basis (Shohamy, 2001a). The standardized test is used as a tool that marginalizes and disfranchises sections of the population and serves to provide a controlling elite with a system for establishing and prolonging their own control (Madaus, 1990; Shohamy, 2001a). Within the social critical analysis of standardized testing, the term method applies to the system through which power is manifest for the marginalization of certain populations and the preservation of power for certain elites. This concept of method in testing is embedded within the analysis of the social system within which the test is utilized. In a sense, social critics have relocated the problem of a method for the human sciences from the cognitive to the social and thus once again relegated the actual problem of a method of interpretation to, at best, a secondary position.

As conceptualized within the current manuscript, the problem of method of interpretation is central to the issue of the employment of standardized tests. The claim made by standardized tests is that they are grounded within a scientific method that allows evaluative numerical comparisons among individuals to be made and it is this claim that stands as the bulwark against the critique of the social critics. In hermeneutic terms, and specifically in relation to human literacy, this is a very strong claim indeed. The history of hermeneutics is based on the understanding that explicit written products are not transparent, self evident, telepathic objects that automatically and effortlessly transmit their meaning to a third party. The basic hermeneutic understanding is that all meaning is the product of a series of historically constituted interpretive procedures (Gadamer, 2003). Rather than self evident meaning, hermeneutics assumes that meaning is interpreted, negotiated, embedded and constructed. Apparently, the method of standardized tests has resolved this age old problem and has a systematic, universally valid, interpretive procedure for understanding human literacy constructs and it is this method that allows the comparison among individuals to be made. If by any chance, the method of standardized testing of human literacy has not resolved this problem of method then
the claims of evaluative comparisons become untenable and, of course, the claims of the social critics against standardized testing become all the stronger.

**A Brief History of the Standardized Language Test in the American Context**

Standardized testing is currently in vogue within the American federal, state and municipal education systems. This current trend is political in essence and can be traced back to the educational policies of the Reagan era. Reagan, when faced with provisional evidence that supposedly revealed widespread difficulties with mainstream literacy abilities, embarked on a program of extensive testing. Within the current ideological climate the proposition that testing is a solution to perceived low literacy levels seems plausible. And yet, even a brief consideration of this idea reveals that it is at best a very indirect route of action. Surely, teaching is the solution to a perceived educational problem with literacy. The explanation for this rather convoluted proposition lies within the ideological realm of the underpinning concept of accountability. The concept of accountability, within the American ideological context, involves a perverse connection between economically defined efficiency and a moral evaluation of individuals and educational institutions. In this proposed ideological framework, educational failure is a moral failing that results from an inappropriate and wasteful use of funds. As constituted within the concept of accountability, testing fulfills a central role. It is through the standardized test that failure is defined and it is through the content and form of the test that controlling elites define the concept of literacy that they are interested in enforcing. This same approach to educational direction has been followed by a series of American administrations and has been further strengthened under the heading of "Leave No Child Behind".

Within the world of language testing the roots of standardized testing can be traced back to the onset of structuralist linguistics and the psychometric tradition (Spolsky, 1978). Essentially, the psychometric approach to testing evolved as a response to localized approaches to language classroom testing at the beginning of the 20th century. Testing was, at that period, seen along hermeneutic lines as an intuitive art rather than a science (Spolsky, 1978). The arrival of structuralist linguistics, which manifested a very significant shift in the definition of language, facilitated the development of psychometric language tests. Structuralist linguistics first transformed language into an object and then compartmentalized the object that it had defined. The crucial component of this shift is the external definition of language as an autonomous object of inquiry. The language form itself can, under this definition, be divorced from the human agent that produced it. Thus the human mind is basically irrelevant for research purposes. It is the underpinning structure found through analysis of the explicit object that is of interest. It is on this basis, the discussion of the linguistic system as an object, that structuralist linguistics viewed itself as a science and not a humanistic art.

Ironically, the definition of the underpinning structure of language although conducted through a textual analysis of the explicit surface form of the object is actually an inference. This creates confusion as to the status of the inference. For many researchers, and more importantly for the wider public, this inference is upgraded to the status of scientific fact and is not understood as an inferential construct. Once this shift has become accepted as an uncontested definition of language a series of subsequent definitional ramifications can be enacted. The first ramification is that once language is an object divorced from the human mind, it is also divorced from any localizing context. Context of production or reception is meaningless because only the language object is of interest and this, according to the structuralist definition, can be studied beyond the context of production. This elevates the linguistic system and devalues the individual human being making the personal
utterance totally insignificant. The second ramification is that as an object linguistic research is centered on underpinning form that is devoid of content. Structuralism's research interests are directed at uncovering the underlying structure of the linguistic object. Communication, knowledge and intent do not figure in this research agenda. As such, form is divorced from content and content is relegated to the role of irrelevancy. The final ramification is that the objectification of language in a decontextualized conceptual setting allows the division of the linguistic system into components. From the structuralist viewpoint the individual components of the linguistic system are the linguistic system. The differentiation between parts and whole is shattered and thus the holistic nature of human linguistic communication is lost.

This structuralist definition of language was fertile ground for the development of psychometric and standardized approaches to testing. By defining language as objectified and compartmentalized linguistic structure that is beyond the human mind, divorced from the locale of production and/or reception and devoid of content, it was possible to construct a decontextualized, abstract and generalized test of segmented linguistic knowledge. The structuralist linguistic project provides a conceptual basis that legitimizes the idea that the test is an autonomous object that represents the individual's linguistic knowledge. As an autonomous object the test designer and the test taker never have to meet, the locale of test taking is irrelevant, the nature and background knowledge of the individual test taker are meaningless, and everything is focused on the test as an object embodying all the relevant information pertaining to the test taker's linguistic knowledge. It is in this conceptual environment that the test-centric approach to testing becomes prominent. The test-centric approach addresses every issue of testing through a manipulation of the object of the test. There is no reality beyond this artificial construct and manipulation of this external object combined with advanced statistical modeling gave the field the false sense of being involved in a scientific endeavor. This approach also gave rise to the major argument used in support of the standardized test. As a decontextualized, autonomous object the standardized test is supposedly objective and adheres to principles of scientific replicability, usually termed within the meta-language of testing, reliability and validity. This publicly projected position of the scientific, objective status of the standardized test was and is a central tenet of the authority these tests seem to enjoy in the public sphere.

In this brief history, it is important to note that the standardized test is a vestige of the past in relation to current developments in testing (McNamara, 2001; Shohamy, 2001b). As in the case of the onset of psychometric testing procedures, new definitions of language and literacy redirected the concept of test design and conceptualization. The trajectory of linguistic and literacy research and theory has moved from an exclusive focus on textual features, to an inclusive redirection of attention to the socially embedded nature of linguistic practice. Whereas once the analysis of a textual product by an authorized expert was considered a legitimate source of the authority of the proposed meaning of the text, current positions pose that meaning needs to be considered in light of the context of meaning construction - including the individual performance of meaning construction, the immediate social context of meaning construction and cultural-historical context of meaning construction. In other words, linguistic meaning and the linguistic system is no longer considered a static, unitary, linguistic entity but rather an interactional, dynamic and virtual product.

In a similar, but delayed, trajectory the field of language testing has moved from an exclusive interest in the test-internal issues of validity and reliability to an inclusive understanding of the social context of testing (Messick, 1994, 1996; Shohamy, 2001a). Language testing has moved from discrete items to integrative tasks and
from tests of static knowledge to an emphasis on performance and process. Understanding of the social context of testing is most clearly developed in the recent work of Shohamy (2001a) and largely comes under the heading of consequential validity. This new perspective of language testing reduces the claim of the scientific objectivity of the standardized test itself as a result of its so-called objective form and internal statistical evaluation, to nothing more than naive self-justification. As every test taker knows, tests have very real consequences and accordingly, the use of the test itself by powerful institutional bodies has very significant social outcomes. The concept of consequential validity involves evaluating these social consequences. One of the more sophisticated analyses of the social consequences of testing is found within the work of Foucault (1979). As quoted in Shohamy (2001a) Foucault proposes that tests constitute what he calls "a normalizing gaze, a surveillance that makes it possible to quantify, classify and punish" (1979, p. 184). In this formulation, and directing a whole host of subsequent social critics of testing, tests are a method by which the socially powerful impose a specific set of norms of behavior on a diverse population. These norms are used to differentiate among groups of people and to reward or punish according to the level of acceptance and use of these norms as evidenced in the testing situation.

Hermeneutics and the Problem of Method in Standardized Testing

Hermeneutics as a field is concerned with the bi-directional move between the internal cognitive world of the individual human consciousness and the external world of sensory expression manifesting socially conventionalized signs. This move from the internal consciousness of thought to the external sign and from the external sign to the internal consciousness of thought is indirect and consists of a stratified series of interpretive procedures that always result in partial results (Grondin, 1994). The core hermeneutic understanding is that meaning is always interpretive, approximate and partial (Grondin, 1994). The major ramification of this core understanding is that methods of research and criteria of investigation for the human sciences are very different than those in the natural sciences. Historically this argument has been developed through the discussion of the problem of "method" for the human sciences. The polemic can be traced back to the beginning of the 19th century (Grondin, 1994). The 19th century juxtaposed Romantic ideologies of the individual and the muse with the successes of scientific methodology. The proven abilities of the scientific method and mathematics at modeling and predicting aspects of the physical world put enormous pressures on the human sciences to propose a similar mode of investigation. Method in the natural sciences was explicitly described and supported by philosophical reflections on the way that science produces knowledge. It was assumed that a similar situation should develop in the human sciences.

This issue of the development of a method for the human sciences was addressed in the 19th century by philosophers such as Schleiermacher and Droysen but it was Wilhelm Dilthey who made a discussion of method the starting point and core of his philosophical work (Grondin, 1994). Essentially, Dilthey aimed to develop a sound philosophical foundation for a method in the human sciences that was clearly differentiated from the method of the natural sciences. Building on Droysen's distinction between the static, sensory world of nature and the changing, intellectually constituted world of history, Dilthey formulated a position that located the human sciences within the inner experiences of the individual. On this basis the object of inquiry for the human sciences, human consciousness, is clearly different from sensory information that can be observed and collected in relation to the natural world. But Dilthey's major claim was that is was not only the object of investigation that differentiated the human sciences from the natural sciences but rather it was the modes of inquiry, the approach to the object, that was crucial (Grondin 1994). Human sciences situated within the inner experiences of the individual mind, posit
understanding and ultimately all validity within the a-priori conditions of consciousness. This formulation of the hermeneutic circle is a predecessor of Husserl's description of the phenomenology of inner experience (Grondin, 1994). In this formulation, the human sciences consist of a movement from external information to internal understanding and experience. This is research into the emergence of the internal world and accordingly self awareness. Hermeneutics becomes the process by which something inward is described in terms of outward signs.

In the 20th century, the understanding of the hermeneutics of interpreting internal thought through external symbolic manifestations was greatly influenced by the concept of Dasein forwarded by Heidegger (Gelven, 1989; Heidegger, 1962). In Heidegger's existential hermeneutics, the Dasein is formulated as an interpretive tendency that is active in every act of understanding. This forestructure of understanding is constituted within a unique existential situation that provides the thematic context and limiting parameters of the specific act of understanding (Gelven, 1989; Heidegger, 1962). This predetermined interpretive framework that is implicitly the manifest self turns every act of understanding into a formulation of the hermeneutic circle. Objectivity becomes an impossibility as every interpretation is preceded and directed by the forestructure of understanding that without reflection remains implicit (Gelven, 1989; Heidegger, 1962). Language poses a particular problem in this respect in that it is simultaneously constructed out of the interpretive nature of the Dasein but hides the phenomenon itself. Thus language seems to objectify but is in actual fact partial, fore-structured and rationalizing (Grondin, 1994). All understanding is situated within the interpretive parameters of fore-understanding.

Gadamer (2003), who posits the problem of method in the human sciences as central to his philosophical agenda, builds upon Heidegger's construction of the hermeneutic circle and the problems of language. For Gadamer (1977, 2003), the principle of historicity is the hermeneutic principle. All understanding is based on the historically constituted forestructure of understanding. Accordingly, the idea of objectivity is unattainable; we cannot escape the historical localization of our existential situation. This basic understanding is developed through the position that every act of meaning production is also constituted within the linguistic structure of the dialogue (Gadamer, 2003). For Gadamer, and in accordance with current conceptualizations of language, language is not an autonomous self explanatory object, but rather a contextualized embedded entity. All utterances can be understood only in relation to the historical localization of the act of production and reception. This position sees understanding as a historical constituted dialogue, with every utterance being related back to the historical context of a question that is being answered (Gadamer, 2003). Methods which objectify language and compartmentalize it, the attempt to apply natural science methods to human sciences, misrepresent the essential nature of understanding as a historically constituted dialogic participation within an extensive tradition of previous questions and answers. Accordingly, the methodology of natural science, with its emphasis on the external, observable and countable, is problematic in that it cannot account for the special inner nature of human cognition and being.

Essentially, the hermeneutic problem is the problem of the strength and closeness of the relationship between the external manifestation of the internal world and the internal world of thought itself. This is a validity problem in that there is no way of knowing if the proposed, interpreted understanding of the internal world does indeed model the internal thought itself. In fact as seen in the discussions of forestructure in the preceding passages, the interpretation of external signs is always historically constituted. So while the method of interpreting signs, modeled on natural science procedures, may be systematic and may provide the same results for similar external manifestations, the question of validity will always be answered in a circular manner.
that involves a regression to an external procedure. The validity question, the essential problem of method in the human sciences, can only be answered by a comparison of the two constructs, the internal and the external. If the comparison is of two external interpretive procedures then validity in relation to the internal construct cannot actually be established. It is for this reason that hermeneutics views method in relation to human constructs as an interpretive art rather than a science and all interpretive procedures as individual, localized, historically constituted, partial and incomplete. The inner thought, word and experience cannot be accessed directly but only interpretively through its external manifestation in conventionalized signs. This is a direct challenge to any approach that will claim to have a universally valid method of interpretation allowing numerical evaluative comparisons among individuals to be made.

Current educational practices in the U.S. and around the world are based to a large extent on large scale standardized testing that are used in a comparative manner with high stakes outcomes for the individual test taker (Hill & Parry, 1994; Shohamy, 2001a). The strength of the standardized test, and one of the main reasons for its continued usage as a tool of educational and social control, is its public image as an objective, unbiased evaluation of knowledge and abilities (Shohamy, 2001a). This claim of objectivity is tied directly to the publicly promoted concept that standardized tests function according to methodologically sound scientific procedures. Over the years and within the test-centric concept of testing, a detailed approach to test construction has been developed. As with other attempts to define an endeavor as scientific, the field of testing has developed a controlling metalanguage and an overriding set of procedures for test development. The metalanguage controls the conceptual understanding of the testing process and the defined procedures provide a way of functioning that constitutes the external definition of a method. On this basis, the concepts of authority, objectivity and scientific method associated with standardized tests are well entrenched in public perception.

The scientific status of the standardized test is wholly dependent on two very significant ontological shifts - the objectification of knowledge and the understanding of the test as an autonomous unified object that can be manipulated in isolation from context. As a very basic definition, a standardized test compares a test taker’s knowledge and ability in relation to a defined set of standards. For this to occur, the first important ontological shift is the move from the knowledge and/or ability of the individual to the definition of explicit, external standards. This shift is crucial in that the whole process of test development rests on this starting point. The overall tendency, in order to ease the test development process, is to define these standards in externally observable, behavioral terms. This is considered the planning stage of the test development process.

This process of explicit, external definition is directly modeled on scientific procedures in the natural sciences and for behaviors that are actually external and observable this process of standards definition is unproblematic. But as pointed out in the discussion of the polemic in hermeneutics over a suitable method for the human sciences, what happens when the knowledge and ability being addressed is within the internal cognitive realm of the individual? What happens, as in the case of the standardized testing of literacy, when the ability being addressed is the ability to construct or produce meaning? These pose a very significant challenge to the planning process because the trait to be tested is actually unobservable directly. The question as to the nature of reading and writing thus becomes crucial. Unsurprisingly, a close look at the methodologies of reading and writing research reveals that all of our knowledge about literacy is actually inferential. Although very rarely stated as such, the field of literacy research is a hermeneutic endeavor to infer the process by which meaning was constructed and/or produced. The status of this knowledge is
partial and fragmentary. In writing circles, with the advent of post process theory, even the idea that an overriding theory of writing could or should be produced is questioned (Kent, 1999). But beyond the ideological arguments, from a purely scientific viewpoint, no methodology exists for the observation of cognitive structures and everything that has been said about literacy is essentially a hermeneutic analysis of external signs in the attempt to recreate or retrace the steps back to the inner thought. As such the state of our knowledge of literacy is truly partial and the very idea of a scientific basis for universal standards of literacy becomes laughable.

To overcome this glaring problem in the definition of standards for literacy, the test designers regress to a particularly unscientific mode of decision making. The standardized test requires the explicit definition of standards. It cannot continue if this process is problematic because all subsequent stages refer back to this planning stage. From a scientific viewpoint this situation needs to be recognized and a suitable alternative research method proposed and adjustments to the type of claim that can be made from such a method. On the basis of partial inferential knowledge, as a researcher you would couch your conclusions in the guarded language of possibility; you would not want to make strong comparative comments about ability. But the process followed by the standardized test designer is to ignore the facts relating to the status of knowledge in relation to literacy and rather apply a definition of standards based on outdated inferential research and socially ordained criteria. Rather than a scientific decision making process, this is the decision making process of the pragmatic politician. In practical terms the standardized test accepts the general structuralist definition of language and adds through the pressures of powerful groups socially defined standards. Thus the standards of the test are generated and defined in explicit behavioral terms. Knowledge and ability have become externalized and objectified. Once the standards are in place they take on a natural reality of their own and the unscientific nature of their invention is forgotten and completely ignored.

Once the standards have been established the second major ontological shift is enacted. As described above, objectivity and scientific method are central to the promoted image of the standardized test. Objectivity is understood within public discourse as complete decontextualization of the test, distance between test taker and test designer, and the equal application of criteria for evaluation. For this to be possible the test itself must be defined as a unitary, autonomous object that can directly transmit knowledge. For standardized testing it is crucial that the test be conceptualized as a self sufficient, epistemologically unproblematic object. The totality of the standardized testing method is based on the principle of the ability of the test to directly reflect the traits being evaluated without any mitigating contextualization. For purposes of standardized testing it is assumed that the test as an object explicitly embodies all the required evidence for an evaluation to be made. Once this ontological shift has been made all subsequent discussions of the test are reduced to manipulations within the test itself. There is no external reality to the test. What is usually considered the scientific method of the test relates to the series of potential manipulations that can be enacted on the test as an object.

It has already been pointed out that current positions view language as a contextualized and embedded entity that requires interpretation. Tests are after all language based and as such the assumed position of self explanatory, unproblematic object is naive and inaccurate in the extreme. Even within the field of testing it is a well established principle that the testing method interferes with the elicitation of data for the evaluation of the test taker (Shohamy, 1984). But within the test-centric approach to testing this is addressed only through the manipulation of the test as an object. The defined standards direct the process of test item and task choice within the context of a unitary finite test object. Authority resides only with the test designers.
and administrators. Factors such as diversity in cultural background, individuality in processing, emotional state, and physical conditions are considered irrelevant. When they are addressed, as a result of social pressure, they are collectivized and once again reduced to a minor textual manipulation in the test itself. From a hermeneutic viewpoint, the idea that construction and production of meaning is the external manifestation of a historically constituted internal consciousness that is part of a dialogue is ignored. For the standardized test to be administered it must have the ontological status of a stable autonomous object. There is no current internal method within the exam that contextualizing feedback addressing diversity is possible. The "other" as a possibility beyond initial conception by the test designer is completely erased as an option. The exam is constructed totally through a non-reflective foreunderstanding of expectations.

The tautological process of the standardized test comes to the fore when the issue of scoring and rating is addressed. Ironically, the ability to accurately evaluate every test taker and the test as a whole is usually considered the strongest aspect of the standardized test. But it is in this component of the testing process that the hermeneutic problem truly becomes acute. Within the metalanguage of testing, drawing upon terms from scientific inquiry, the issue is the procedure for ensuring reliability and validity of the test. In broad terms the scoring of standardized tests is directed by an a-priori decision as to the correct type of response. This decision is dictated by the original definition of the standards, hence the basic tautology of this process. This situation is true for closed items in which the answer is a choice made by the test taker or in the situation in which an open ended response is given by the test taker and a rating scale for scoring developed. In both cases an explicit response is rated in terms of the explicitly defined external standards. The response is taken as a direct representation of the internal world of the individual. The defined correct response and the rating scale are not considered as hermeneutic decisions and accordingly their partial interpretive nature is not addressed. Although within the test procedure the response is taken as explicit and self meaningful, from a hermeneutic viewpoint the defined answer and the rating scale are interpretations.

A lot of effort has been put into the procedure for the systematic development of responses and rating scales. The development of these responses and rating scales is usually conducted through analyses of large scale responses and the usage of the rating scale in multiple trials of a test. The aim of these analyses is to look at the systematicity of the items, tasks and rating scales in relation to the proposed aims and objectives of the test. The data set is the responses to the test as explicitly presented in the answers given in the test. Statistical analyses look at the distribution of the scores, the ability of items to discriminate among test takers, the consistency and accuracy of the rating and the difficulty of the items. These analyses and others allow the test designer to evaluate the product that has been produced. But they all function within the closed data set of the responses given. They do not have the option of addressing information beyond this data set. From a hermeneutic viewpoint the data set that is being analyzed statistically is partial and in fact the most crucial data, the internal world of the test taker, is not present at all. These statistical tests are very good at making sure that the test as a unitary object is reliable and that the scoring of the test is accurate in relation to the proposed standards. All tests of reliability compare two explicit situations of response, whether it is two parallel versions of a test, the relationship among test items or the usage of a rating scale by different raters or the same rater on different occasions. Statistical analyses are very useful in the situation of comparing explicit responses and adjustments can be made to the test or rating scale to make sure that reliability is high.

Validity however is a much more complex problem and it is here that the whole system of standardized testing as a scientific endeavor is seen to be an illusion.
Validity is the process in which the test is evaluated for its ability to measure the trait it was designed to measure. In the case of standardized test of literacy the traits to be measured are reading and writing. As discussed above, both of these are internal processes to which we only have indirect interpretive access through the external manifestation of conventionalized signs. Like reliability, validity is essentially a comparative process in which two manifestations of the same thing are compared and closeness calculated. The hermeneutic problem of method defined above and applied to the issue of standardized testing states that for the standardized test of literacy to be valid it must have a procedure that allows the comparison of the defined test with the internal construct of the individual. For the scientific status that the standardized test requires for itself, this procedure must also be systematic and universally valid supported by statistical analyses showing very little divergence between the internal and external construct. Unfortunately such test is impossible because of the lack of access to the internal cognitive realm of the individual. All that can be done, and what is done within the world of standardized testing is to regress to a comparison of external inferential constructs of reading and writing. The procedure is the comparison of the test in relation to the proposed standards and to other explicit tests or grades. While this procedure does allow a score to be calculated and statements made about the validity of the test, this does not in any way solve the basic hermeneutic problem. The presence of statistics does not make something scientific and the comparison of two explicit manifestations of an exam or an exam to later scores does not make a standardized test of literacy valid. At this time the standardized test of literacy cannot be scientifically shown to be a valid measure of literacy knowledge or ability. According to the criteria set by the standardized testing bodies for themselves, a scientifically sound standardized test must be evaluated for its reliability and validity. In the case of standardized tests of literacy, only reliability can be established. Validity is still an open issue. Accordingly, the status of the test must be adjusted to take the hermeneutic problem of method into account.

This problem of validity has very serious ramifications for the test reporting process. In a situation of limited validity the status of the reported results must be stated in very careful terms. Since the validity of the test is in question and it is not clear what exactly is being measured, it is very unwise to make strong claims about the results themselves. As stated in the previous paragraph the presence of a percentage grade that is reliable does not translate into the ability to conduct comparisons among test takers with the assumption that the test represents the literacy abilities of the individual. Within the current educational climate in the U.S. and around the world, serious decisions in relation to funding, access to higher education and progression through grades are based on standardized tests. The test results are reported as if the issue of validity has been solved. In actual fact this is a false claim and as shown in this article the question of validity cannot be ignored or covered by a thin veneer of meaningless statistical comparisons to other tests or the invented standards themselves.

As pointed out by Spolsky (1994) developments in testing procedure have managed to produce very reliable tests. But the real question is in relation to the validity of these tests (Messick, 1994, 1996; Spolsky, 1994). In simple terms the question becomes what is it that the tests test. The hermeneutic problem of method as defined in this paper interacts with the proposed system of standardized testing of literacy at three key junctures: the definition of standards, the contextualization of test taking, and the scoring and reporting of test responses. On the issue of standards for literacy, the hermeneutic problem of method manifests itself in the problematicity of objectifying literacy knowledge. The absence of a scientific method for defining literacy as an internal construct has left the field of literacy research as an exercise in hermeneutic discovery. The ramifications of this are that the status of our knowledge
is partial and thus it should not be universalized and objectified. On this basis
standards are to be considered at best inferential and partial constructs rather than
scientific factual entities. On the issue of contextualization, the hermeneutic problem
of method specifies that context is the major component of defining any form of
hermeneutic. When addressing the understanding of explicit manifestations of
internal cognitive constructs an individualized, localized and contextualized
hermeneutic is being engaged. Language is not self meaningful and self explanatory.
On this basis the status of the decontextualized standardized test is that it provides
only partial information that is relevant for evaluating literacy abilities and that this
partial knowledge is itself the result of a hermeneutic process. On the issue of
scoring and reporting tests, the hermeneutic problem of method specifies that there
is at this present time no scientific way of evaluating the validity of the proposed
standardized tests of literacy. Accordingly, although these tests can be evaluated for
their internal and external reliability, the status of what they are actually measuring,
their validity is an open question. On this basis the presentation of numerical scores
as unproblematic entities that can be used for conducting comparisons and
representing literacy abilities is untenable. The scores produced in these tests are
artificial, interpretive constructs that lack an evaluation of validity. In this sense they
need to be carefully reported and couched in qualifying language that accurately
reports the partial, interpretive nature of the score given.

Final Remarks

The hermeneutic problem of method as it interacts with the issue of standardized
testing poses an unsolvable dilemma for standardized tests of literacy. Essentially as
discussed in the article, the validity of standardized literacy tests cannot be
established. What usually passes as validity is the comparison of the test to
externally defined standards (usually based on outdated research and socially
defined "desirable" outcomes) and other external grades or scores. While these
attempts to establish validity are backed by an impressive discourse of scientific
procedure and statistical reliability, the problem cannot be solved by these measures.
Only a comparison between internal and external constructs will establish the validity
of the literacy test. Thus the discourse and associated scientific and statistical
procedures become nothing more than a rationalization and a rhetorical ploy of
authority to quieten and overcome opposition and impose social control. As
described by Shohamy (2001a) standardized tests are the first choice of educational
policy makers and politicians in order to direct educational content and enact social
control. It is very unlikely that the problem of validity as described in this paper will
change this situation. However, for literacy educators, literacy researchers, school
administrators and exam participants it is important to understand that the
standardized test of literacy is a problematic tool that cannot be said to validly
represent the literacy knowledge or ability of the individual test taker. This
understanding may contribute to a grass roots reconceptualization of the nature of
the standardized tests of literacy and eventually undercut the power of these tests in
directing and controlling our lives.

References

York: Continuum Publishing Company.

California University Press.

Haven: Yale University Press.


