DOCUMENT RESUME

ED 454 270 TM 032 879

AUTHOR Munby, Hugh
TITLE Educational Research as Disciplined Inquiry: Examining the Facets of Rigor in Our Work.
SPONS AGENCY Social Sciences and Humanities Research Council of Canada, Ottawa (Ontario).
PUB DATE 2001-03-00
NOTE 15p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (St. Louis, MO, March 25-28, 2001). From the research program, "Co-op Education and Workplace Learning" (Hugh Monby, Nancy Hutchinson, and Peter Chin).
PUB TYPE Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Research; Ethics; Models; *Qualitative Research; *Reliability; *Research Methodology; Rhetoric; *Validity
IDENTIFIERS *Professionalism; *Rigor (Evaluation)

ABSTRACT

This paper explores how facets of the concept "rigor" might be applied to questions about the validity and reliability of research independently of the research modes. The focus of the critical lens could then be on how to assess the contribution of various forms of research rather than on the "paradigm wars" and arguments about various research modes. The paper opens with a brief look at theoretical frameworks that acknowledge the legitimacy of different forms or modes of inquiry and allow a more direct focus on rigor within different forms. The discussion of rigor presents a recent history of the concepts of reliability and validity that tracks changes in meaning, followed by an illustration of how these concepts work together to provide a sense of rigor. It is suggested that rigor needs to account for the application or use of research, opening the way for looking at several aspects of rigor including ethics, professionalism, and rhetoric. The discussion of these issues is framed by quotations from "Under Which Lyre: A Reactionary Tract for the Times" by W. H. Auden (1946). (Contains 40 references and 7 figures.) (SLD)

EDUCATIONAL RESEARCH AS DISCIPLINED INQUIRY: EXAMINING THE FACETS OF RIGOR IN OUR WORK [1]

Hugh Munby

Faculty of Education
Queen's University
Kingston, Ontario K7L 3N6

Introduction

Thou shalt not sit with statisticians nor commit a social science

These words are from "Under Which Lyre" by the British poet Wystan Auden (1907-1973) in his Phi Delta Kappa Poem at Harvard 1946, shortly after the end of the Second World War. Social commentaries have taught me to think that those were "heady" days: the universities and colleges were alive with veterans, and the continent was riding a wave of euphoria and optimism not experienced since before the Great Depression. In "Under Which Lyre," Auden was speaking especially to the war veterans: "Thou shalt not sit with statisticians" is among Auden's "Hermetic Decalogue" or ten commandments for university students from "precocious Hermes." Those who know that most of my research over the last 30 years has been qualitative may be excused for supposing that I have selected the above words because they appear to favor one approach to research over another. Reading more of his poem, we find that this perhaps is not the case:

Thou shalt not write thy doctor's thesis on education

Auden's poem, subtitled "A Reactionary Tract for the Times," rails against elements of a mass tertiary education and celebrates both intellectual and sensual exploration. His concerns about freshmen (yes, it was 1946) students at Harvard, Princeton, and Yale remind me of teacher education students who approach their programs with the view that there is a finite set of clear rules of procedure that, when followed, lead to good teaching. And of course, research students mirror this behavior if they slavishly follow detailed steps for ensuring that their quasi-experimental designs meet criteria for validity. Neither are students of qualitative research immune from the infection of checklists. As Webb and Glesne (1992) noted, "Some students assume that a qualitative research class will provide procedures that, if followed faithfully, will produce warranted research results" (p. 775).

But what could "warranted research results" mean? I have just suggested that there is more to "warranted research results" than having the researcher satisfy familiar checklists, such as those in the five editions of McMillan and Schumacher's (1984-2000) *Research in Education: A Conceptual Introduction*—a text that I continue to use once or twice a year when I teach our introductory research methods course. In this course, I try to have students understand that, ultimately, our research is a human enterprise and that its worth is more than its trustworthiness. So although "warranted research results" probably has something to do with trustworthiness, and with concepts like reliability and validity, I suspect that there is more. And that is what this paper is about.

---

[1]Invited address, annual meeting of the National Association for Research on Science Teaching, St. Louis, MO, March 2001. This paper is from the research program "Co-op Education and Workplace Learning" (Hugh Munby, Nancy Hutchinson, and Peter Chin, Principal Investigators) funded by the Social Sciences and Humanities Research Council of Canada. [munbyh@educ.queensu.ca]

The purpose of this paper is to explore how facets of the concept "rigor" might be applied to questions about the validity, and reliability of research independently of the research modes. In this way, the focus of our critical lens can be on how to assess the contribution of different forms of research rather than on the somewhat tiresome "paradigm wars" and their overworked arguments about various research modes.

The overall approach to the paper is to open with a brief look at theoretical frameworks. This acknowledges the legitimacy of different forms or modes of inquiry and allows us to focus more directly upon rigor within different forms. The venture into rigor begins with a recent history of the concepts reliability and validity that tracks changes in meaning. This is followed by illustrations of how the concepts work together to provide a sense of rigor. I then move to showing that rigor needs to account for application or use of research, and this opens the way for looking at several facets of rigor including ethics, professionalism, and rhetoric.

Theoretical Frameworks

The examples I have of theoretical frameworks are of those that operate at a relatively high level in the interpretation of data. The examples suggest that issues of rigor lie beyond debates about qualitative and quantitative research. One example is from the discipline of history. I was a doctoral student in the late 1960's—those too were heady days, with Canada's universities alive with social protest and Mary Jane. As eager students of education, we read such texts as *Growing Up Absurd* (Goodman, 1960), *How Children Fail* (Holt, 1964), *36 Children* (Kohl, 1967), *Life in Classrooms* (Jackson, 1968). Also among my readings was the debate about the function of history with Hempel's (1968) work on the side of proposing explanations for human behaviors as deductions from general laws and Dray (1957) and others arguing for the unique character of historical explanation lodged within a singular context. I recently encountered this conversation continued in letters between proponents and exponents of quantification in Aydelotte's (1971) *Quantification in History,* and I recalled the importance of distinguishing between two kinds of argument in this context. The first kind is about what overarching approach is proper and should be taken to research. The debate between Hempel and Dray typifies this kind of argument. The second kind of argument is about the quality of the research itself. This is where reliability and validity find employment, and it is where I think rigor resides.

My second example illustrates how our views about what frameworks are proper are modulated over time. The following account of an approach to English literature is from a recent book by Davis, Sumara, and Luce-Kapler (2000):

> at the start of the 20th century, there was a powerful movement among literary scholars that came together around the belief that literary texts could and should be closely read for their exact meanings. That is, according to this movement, literary text should be considered in the same category as non-literary texts. Working from the premis that close analyses of text construction could yield accurate and consistent insights about the author's true intentions, these scholars labored to develop particular methods that were modeled according to those used by mathematicians and scientists of the time. (pp. 231-232)

Arguments against this framing of literary criticism might point out that "novels, poems, and other texts that were deliberately written to maximize interpretive possibilities for readers came to be read in ways that foreclosed on those possibilities" (p. 232). For example, we would lose almost all the richness of contrast that Auden develops in the stanza:

Encamped upon the college plain
Raw veterans already train
    As freshman forces;
Instructors with sarcastic tongue
Shepherd the battle-weary young
    Through basic courses.

The example from a literary perspective allows me to indicate how my understanding of science seems to have changed as a consequence of the work of scholars, some from disciplines other than science. Conant's (1957) richly informative and literary accounts of experimental science from an historical perspective opened my eyes to the fundamental character of science. What I saw was quite different from what I had been taught at school and university, yet it enriched what I had been taught. So it is that I find Bruner's (1996) claim, "The process of science making is narrative" (p. 126), unsurprising as he describes the difference between finished science and the "lively processes" (p. 127) of science making. In this light, we can imagine how we have become accepting of alternative approaches to educational research. Similarly, we should not be surprised to find sections on bivariate and multivariate statistics in LeCompte and Schensul's (1999) *Analyzing and Interpreting Ethnographic Data*. It is almost as if "paradigm wars" have been transmuted into paradigm rapprochements in which different viewpoints let us see better the human condition within our research. In turn, this suggests to me that there is more to assessing research than what is conveyed by reliability and validity.

Reliability and Validity: Meanings from Reputable Sources

Sometimes I hear suggestions that qualitative researchers should avoid using terms that might appear too closely allied to those used in quantitative research. "Subjects" was one of these—it is frequently replaced by "participants" in recognition that much qualitative research is founded upon commitments to individual constructions of realities. The terms "reliability" and "validity" appear on the restricted list too. A recent example is the following:

> We feel that these terms are misappropriated from a more positivist paradigm of research . . . and that some (research teams) are misguided in their striving for concepts such as interrater reliability. (Barry et al., 1999, p. 27)

And this from Janesick's (1994) "Dance of Qualitative Research Design":

> Implicit in the member-check directive however, is the psychometric assumption that the trinity of validity, generalizability, and reliability all terms from the quantitative paradigm are to be adhered to in research. I think it is time to question the trinity. (p. 216)

In my view, it is also time to question the lineage. So before I probe current meanings for terms like reliability and validity, perhaps I might set the record straight by reporting on a search directed at the *Oxford English Dictionary* (Simpson & Weiner, 1989) for words like reliable, reliability, valid, and validity. The *Oxford English Dictionary* was itself a project of immense proportions involving significant care and rigor for more than 70 years (Winchester, 1998). The aim was to fix meaning in the English language, but not in the sense that the French language is fixed by l'Académie Française. The latter attempts to control usage, while the *Oxford English Dictionary*'s guiding principle "is its rigorous dependence on gathering quotations from published or otherwise recorded uses of English and using them to illustrate the use of the sense of every single word in the language" (p. 25).[2]

As one may have guessed, the words on our forbidden list are old. The earliest record of "reliable" is from the 1569 Registry of the Privy Council of Scotland:

> Thair deliverance..and jugement to be als reliabill..as gif the samyn were gevin..be the Lordis of Sessioun.

[2] The first editor of the *Oxford English Dictionary*, Dr. James Murray, began systemically compiling the dictionary from the words and quotations sent to him by readers. A major contributor to this endeavor was Dr. William Chester Minor, who contributed some 10,000 slips upon which were recorded uses of words. Minor, previously a surgeon captain in the US Army, was "detained at Her Majesty's pleasure" in Broadmoor Asylum for the Criminally Insane having been found not guilty by virtue of insanity in the case of a shooting incident in Lambeth, south London. The incident occurred near St. Mary of Bethlehem Hospital for the insane, from which the word "bedlam" is derived. Oddly, the murder weapon was a Colt, though perhaps not Old Reliable—one of the very few instances of this word being used as a noun, according to the *Oxford English Dictionary*!

The use of "reliable" in statistics to refer to concordant results comes over 300 years later, in 1892, from volume 17 of the journal *Analyst*. And Coleridge uses "reliability" in 1816, some 90 years before the term appears in serials like the *American Journal of Psychology*, where it appears in volume 15 in 1904, and almost a century earlier than when Spearman wrote the following in the volume 3 of the *British Journal of Psychology* in 1910:

> A very convenient conception is that of the "reliability coefficient" of any system of measurements for any character. By this is meant the coefficient between one half and the other half of several measurements of the same thing. (Cited in the *Oxford English Dictionary*)

Examples of Validity

Terms like "valid" and "validity" have an equally venerable past. Here, the earliest examples are to law. Scotland again, this time 1571: Seing his said tak is valide and sufficient in the self.

And soon after are examples showing "valid" used of arguments, proofs and assertions, as is Bentley's 1692, "He may admit of those arguments as valid and conclusive."

Similarly, "Two or three daies after, he began to discuss with him the validitie of his maryage" is from *Life Fisher* circa 1550. And in 1581, we could have read, "Of no greater valydyty is that argument lykewyse which they rake out of Augustines wordes" (J. Bell *Haddon's Answ. Osorius*).

Finally, I derive a certain comfort from this passage in 1881: "A generalisation obtained from one book would be fairly valid for all the rest." These terms are not simply "terms from the quantitative paradigm."

Contemporary Confusion

In the quantitative social sciences, reliability is connected with the reproducibility of results, and it has come to be associated with agreement across cases and observations. Most particularly, the term becomes a property of instruments for mental measurement (Gould & Kolb, 1964), although reliability or stability of data can be concerned with the reliability of the observer, the coder, and the analyst. And this sense seems to coincide with how the term may appear in qualitative research, especially in ethnographic work. But as we pursue this, so the matter becomes complex.

For example, studies themselves, experimental or descriptive, can be judged for reliability. Goetz and LeCompte (1984), state that "Reliability refers to the extent to which studies can be replicated" (p. 211), and so "external reliability addresses the issue of whether independent researchers would discover the same phenomena or would generate the same constructs in the same or similar settings" (p. 210) while internal reliability "refers to the degree to which other researchers, given a set of previously generated constructs, would match them with data in the same way as did the original researcher" (p. 210). I find this becomes confusing when the authors attempt crisp definitions of validity:

> Internal validity refers to the extent to which scientific observations and measurements are authentic representations of some reality; external validity refers to the degree to which such representations can be compared legitimately across groups. (p. 210)

These ideas cohere well with the entries in *A Dictionary of the Social Sciences*, although we might be justified in being confused by the idea of "corroboration of one's data" (Gould & Kolb, 1964, p. 742) because it resembles ideas of reliability. And students of research methods can be excused their confusion too. The same sense of corroboration exists in McMillan and Schumacher's (1997) account of internal validity: "Validity of qualitative designs is the degree to which the interpretations and conceptual categories have mutual meanings between the participants and the researcher" (McMillan & Schumacher, 1997, p. 404). And their definition of external validity differs from that Goetz and LeCompte's (1984), but they put the concept in terms of comparability or extension, even usefulness of a study, "the degree to which the research design is adequately described so that researchers may

use the study to extend the findings to other studies" (p. 411). And the student might be further frustrated in a quest for clarity if he or she encountered texts in which conventional terms like "internal validity," "external validity," "reliability" and "objectivity," are replaced respectively by "credibility," "transferability," "dependability," and "confirmability" (Hoepfl, 2000, p. 9). Other versions of the terms and their meanings exist, as in Moschkovich and Brenner (2000, p. 479) for example.

Precise definition, it seems, eludes our grasp. To approach the idea of reliability, Bogdan and Biklen (1998) ask "Will two researchers independently studying the same setting or subjects come up with the same findings?" (p. 35), and state:

> This question is related to the quantitative researchers' word *reliability*. Among certain research approaches, the expectation exists that there will be consistency in results of observations made by different researchers or by the same researcher over time. Qualitative researchers do not exactly share this expectation....In qualitative studies, researchers are concerned with the accuracy and comprehensiveness of their data. Qualitative researchers tend to view reliability as a fit between what they record as data and what actually occurs in the setting under study, rather than the literal consistency across different observations, (pp. 35-36)

Although I have thought about statements like this for some time, I still fail to detect a clean distinction here between reliability and validity. What "actually occurs in the setting" is unknowable except as a construction of participants or observers, and so the issue of reliability seems to depend upon validity to some extent. My confusion is not eased when I consider the large number of terms used in qualitative research for expressing validity and reliability, nor when I see that meanings tend to be somewhat mobile. My experience is that this state of affairs is unwelcome to graduate students in research courses, but more importantly it may be leading me and them in the wrong direction. The enterprise of discussing validity and reliability from varying viewpoints can too quickly involve us in debates about word usage. This can distract us from seeing that research at its most fundamental is an *argument* that leads us through purpose, related literature, data, and analysis to a specific point. This rather oversimplifies, and it omits ideas about the basic frameworks used and about the devices that permit moves from data to analysis.[3]

I believe we can make progress if we focus on argument itself. Certainly, the confusion about validity in mental measurement has profited from a similar switch in vantage point. Views of test validity have changed markedly over the last half century. In the 1950's, validity was construed in four separate ways: content validity, predictive validity, concurrent validity, and construct validity (APA, 1954). And 35 years later, we find Messick (1989) developing his position that these are not separate forms but are evidence for the one form: construct validity. Validity in this frame is an *argument*. A similar commitment to argument is evident in Mischler's (1990) position. He draws on his experience in narrative research to show that validity is less important than the process of validation. He argues that "validation is the social construction of a discourse through which the results of a study come to be viewed as trustworthy for other investigators to rely upon in their own work" (p. 426). The attention to process suggests the promise of looking at argument to get a fuller sense of what is involved in the concepts of reliability and validity and how they might contribute to rigor and to showing the human character of our research.

Depicting Validity and Reliability within Argument

In developing his theory of physical reality, Henry Margenau (1950), formerly professor of Natural Philosophy and Physics at Yale, offers a model (Figure 1) to show the distinction between the protocol data we receive from Nature and the constructs we derive in attempts to describe and then explain. The protocol data are represented on a plane because they have no analytic depth, in contrast to the concepts or constructs in the C-field. Lines between constructs and Nature's plane are intended to suggest measurement, and lines among constructs depict the interrelatedness of constructs within a theory or theoretical system.

[3] Roberts (1982) uses diagrams from Toulmin's ideas about the structure of argument to illustrate such features, and he shows that although different empirical research modes have different features, like metaphysical premises and warrants, they share a common commitment to argument based on data.

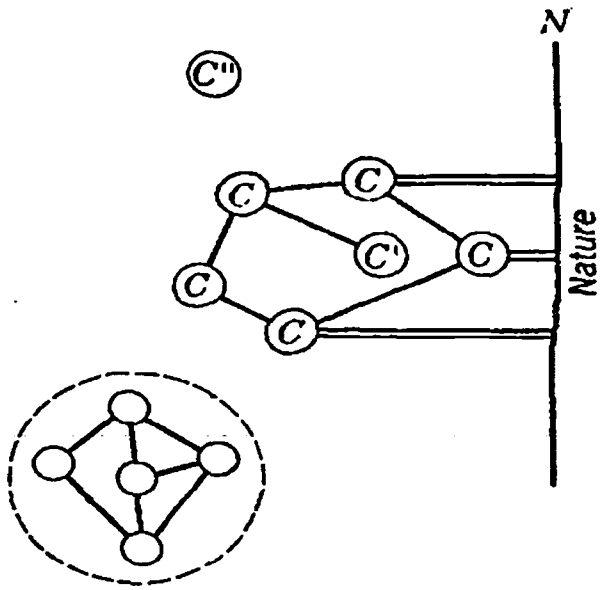

Figure 1. Henry Margenau's initial representation of physical reality.[4]

In a later work (Margenau, 1972), the model is rotated through 90° as in Figure 2, giving a more familiar picture of the vertical relationship between data and constructs, with more encompassing theories being more distant from the protocol plane. Over the years, I have found it helpful to use adaptations of this model to show relationships between ideas of reliability and validity, to explain something about sampling (especially the difference between target and sample population) and to demonstrate the idea that research is an argument.

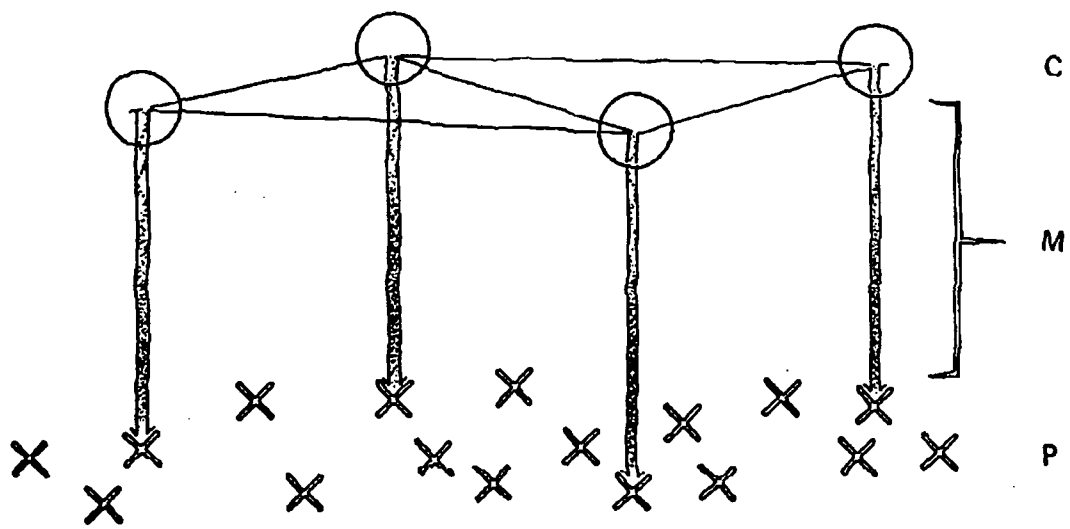

Figure 2. Henry Margenau's representation of physical reality rotated through 90°.

[4] C" is not connected to data or to other constructs. The cluster of constructs at the bottom left have only logical relationships with one another.

Figure 3 represents a study in which a limited amount of data is used to create a set of constructs or a theory. An example from the research on co-operative (school-to-work) educational programs we have conducted might be the understanding that, in a veterinarian clinic, a student's learning of how to prepare a sterile pack is cued by having the student imagine the sequence of the surgical procedure in which the pack is used. The figure shows clearly that the claim is limited in its range. Indeed, an attempt to apply it to other situations, to other data, is risky from a purely structural point of view: a horizontal line drawn from the construct box would disturb the equilibrium and the structure would tip.

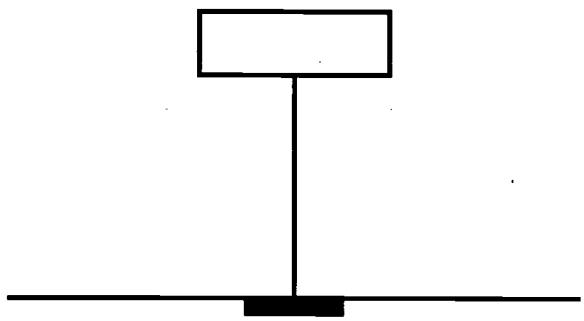

Figure 3. Construction of theory from a limited set of data.

A further feature of the structure is how it depicts the idea of validity. The constructs that, in this case, give us the theory about student learning, are supported by the arguments made from the data. Validity seems to operate vertically here, as shown in Figure 4.

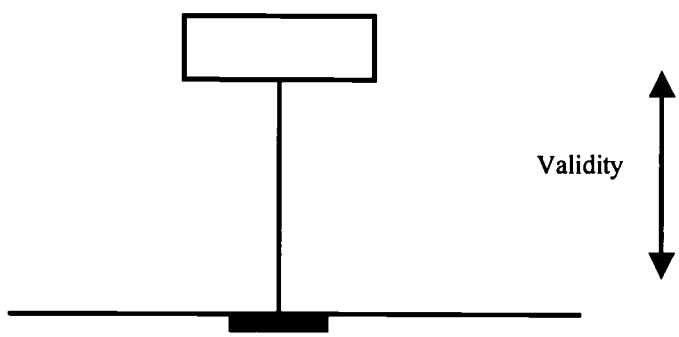

Figure 4. Construction showing vertical nature of validity

The somewhat precarious situation of the constructs we have been using in our example can be stabilized in several ways. One of these is represented in Figure 5. Here, the research team gathers more data. As we have seen earlier in the discussion of reliability, corroboration with additional data (sometimes from different modes of data collection) enhances the reliability of a study. In this case, the base upon which the arguments rest is extended so that reliability, as agreement, is represented horizontally. As more data are added to the research team's files, so the arguments are elaborated as Figure 5 suggests, and the resulting structure is more stable than that represented in Figure 3.

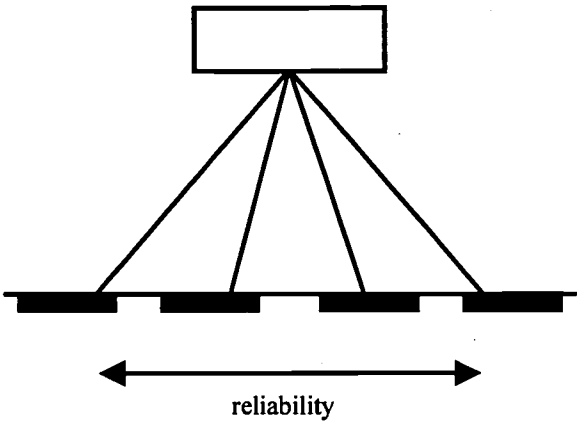

Figure 5. Construct is strengthened by use of more (and corroborating) data

In our research, we often work individually on copies of the data and then bring our interpretations and arguments together for comparison and discussion. The resulting research might look like Figure 6, in which the directional relationship between reliability and validity are seen: together, the two create a stable structure. Another stable structure results when independent researchers working with similar data create similar constructs, as suggested by Figure 7 in which many lines would be drawn linking the work of the two researchers as they inspect each other's arguments.

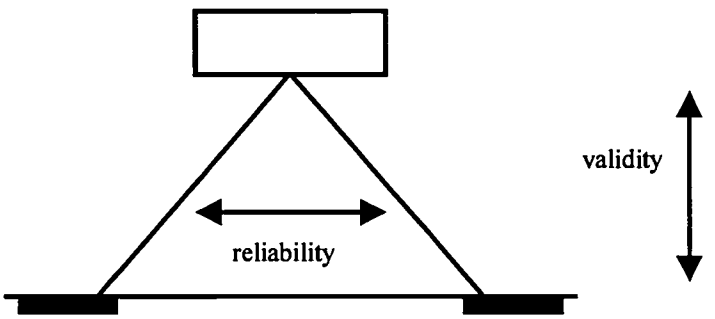

Figure 6. Researchers agree on the validity of the analysis

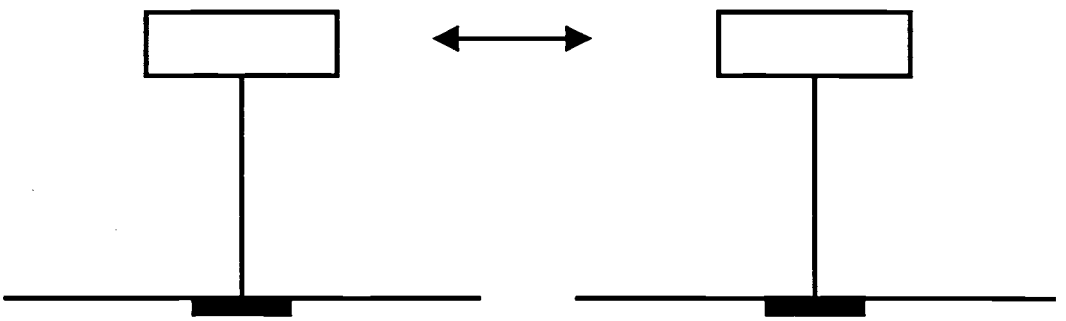

Figure 7. Independent researchers create similar constructs.

I have found figures like these to be helpful in teaching about research. The figures invite us to consider the overall structure of the research argument and its strength, and so offer a context for understanding checklists of threats to validity or steps to increase reliability. Also, by representing validity and reliability as vectors, the figures show something of how the two work together to fashion the idea of stability in our research arguments.

But I am far from comfortable that validity and reliability tell us all that should be told about the quality of research in education. The concepts seem necessary but not sufficient to a full account. What discussions of trustworthiness, credibility, reliability, validity seem to lack is the sense that *research has a purpose*. Here I am not referring to the purpose we might find in a section called "Statement of Purpose." Rather, I am interested in what we think research that we do is *for*. Again, the standard accounts are a little deceptive. McMillan and Schumacher (2000), in their latest edition, announce, "Research advances knowledge and improves practice" (p. 17). In fairness they then consider several different uses of research and develop these into basic (pure or fundamental), applied, and evaluative functions of research. None of this is contestable, I suppose, it's just incomplete. For example, it fails to acknowledge that, among other things, research is to persuade. In the next section, I explore aspects of rigor by considering research purposes. All this is to suggest that conventional tools like reliability and validity are simply not up to the task of portraying what needs to be said about the quality and usefulness of research.

Looking for Rigor in the Purpose of Research

What the diagrams seem to miss, and what I think we need to show to our research students is how the constructs we build get transported into arenas of professional practice, into the settings in which they can be used. My experience is that this transportation is not always successful. There seems to be a membrane between the construct field and arenas of practice. Presumably, if our constructs were objective, in some sense, the membrane would be easy to cross. But that option is no longer available to us.

> When educational researchers no longer see the possibility of objectivity as a life option, one reaction has been to focus on their subjectivity, to worry about it, and to turn it into a set of methodological concerns. For a number of researchers, anxiety about how to be objective as possible has been translated into anxiety about how to manage subjectivity as rigorously as possible. (Heshusius, 1994, p. 15)

There are several ways in which researchers have reacted to the challenge. Heshusius, for example, advocates a methodology of participatory consciousness. My approach is rather different, indeed it starts from a different place. Basically, I do not think I have ever been wedded to objectivity itself because of the character of the knowledge produced by educational research, and because of its point. Indeed, I find a focus on point or purpose particularly helpful in describing something of the range of debates that we should enter when we consider rigor seriously. As I show below, these debates should include issues of ethics, professionalism, and rhetoric.

Ethics and Rigor

I became concerned about these issues when I was asked to write on the significance of qualitative research (Munby, 1983).

> The unquestionable purpose of the enterprise of educational research is the improvement of education. Generally, setting aside research that is more conceptual in nature, it is easy to see that quantitative and qualitative investigations of school events are designed to improve what occurs in educational institutions. While the foci of this work may run from research on classroom learning to research on curriculum change, the ultimate change held as the end-in-view has to be change in teaching practice, because what really counts is the chalkface, curriculum-in-use facet of the endeavor. Here, though, there is an implicit assumption that teaching is the sort of activity that *can* be changed. The corollary is that teachers can be changed. Of course, accompanying these

assertions is the driving belief that research is worthwhile because teachers need to be changed. (p. 424)

For me, it is important to capture the idea that research activity begins with a normative premise. It has never been sufficient to justify research in terms of knowledge for its own sake. Indeed, I have come to think that all propositional knowledge is in the service of action (Munby, Chin, Hutchinson, 2000), and action is clearly normative. In part, the normative nature of our research is reflected in our insistence that there be a rationale for the work. My hope is that the insistence carries into explicit statements about the value premises underlying the proposed work. Without these, I think the research would fall somewhat short of meeting a standard of rigor, and that standard is patently not an objective one when value premises are at issue.

Professionalism and Rigor

Earlier, I argued the impossibility of smoothly moving from generalizable research results to changing teachers principally because the particular circumstances of a teacher's action will be different from those in which the research was conducted (Munby, 1983). In quantitative research, we recognize this issue as a version of the separations among target population, sample population, and sample. In qualitative research we recognize this issue as part of the character of the research too: there is no pretense to generalizability. Here the membrane between research knowledge and professional practice is more than a matter of logic though. Professional assimilation in the field also plays a part.

The expectation that our research might be immediately directed toward teachers suggests that we look carefully at the concept of professional autonomy, "because the latter is imbued with understandings about independent and thoughtful action" (Munby, 1983, p. 426).

Discussions about professionalism are almost as wide-ranging as definitions of what professionalism entails: in the latter we find (a) the contrast between doing something for pay and doing something free, (b) the idea that being professional involves a distancing or detachment (as in calling penalties while refereeing), (c) the suggestion that a degree of proficiency if not excellence has been achieved, and (d) the social distinction among classes that might be reflected in discriminating among occupations, vocations, and careers (Soder, 1990). Some of these discussions tend to agree that professionalism is bound to the idea of a professional knowledge base (e.g., Fenstermacher, 1990).

Colleagues and I have argued that "the essence of professionalism is professional action" (Munby, Russell, & Martin, in press) and that teaching should be in the best interests of the clients and thus based upon the best available knowledge. But as shown in our chapter in the fourth edition of the *Handbook of Research on Teaching*, the character of teachers' knowledge is the subject of debate and conflicting theoretical viewpoints. This makes the transition of research knowledge into professional practice highly complex. And in turn, questions about the quality and value of educational research automatically get extended beyond the simple language of reliability and validity. A sense of rigor is called for that honors both the moral premises of research purpose and the integrity of professional knowledge and judgment, without violating the professionalism of the educator.

Rigor and the Researcher's Professionalism

The *Oxford English Dictionary* reports many senses of "rigor" from strict application of the law, through hardness of heart, to strict accuracy and severe exactitude, a phrase that seems to refer to lexicography itself. Also we have seen how rigor gets entwined with professionalism, so it is fitting to turn the lens on the professional actions of the researcher himself or herself and to ask how rigor gets played out in that arena.

I doubt that I am alone in wondering along with graduate students at the quantity of research decisions we face that are not strictly guided by anything epistemological. Questions like, "How many participants should I really have? "Are eight interviews enough?" "Should I attempt another administration of the test or simply go for a split-half assessment of consistency?" As a graduate supervisor, I often find myself saying, "This is just a masters thesis, not a career" so truncating research for purely practical purposes. Of course, the section titled "Limitations" always

accounts for how practicality may compromise rigor, but somehow we miss saying that we, as professional researchers, do this all the time. Among my favorite quiet compromises are the following.

The first is the rule of thumb we seem to have developed for the reliability of instruments. Noting that reliability is a function of the nature of the trait (construct) being measured, McMillan and Schumacher (2000) state, "a reliability of .80 or above is generally accepted for achievement variables, whereas estimates of .70 may be acceptable for measuring personality traits" (p. 249). I spent an early part of my career wondering about the reliability and validity of attitude measures. I won't go into details here, but it is worth observing that reliability in this sense has become something of a rhetorical device rather than an epistemological one.

The second example is the threat to internal validity of treatment replications: "In an experiment, the treatment is supposed to be repeated so that each of the members of one group receives the same treatment separately and independently of the other members of the group" (McMillan & Schumacher, 2000, p. 191). If an instructional treatment is conducted once in one class, then the class is like one subject. The sample size is the number of treatments, not the number of subjects. The threat of treatment replications refers to instances when the reported number of subjects is not the same as number of treatments.

I wonder if we are deceiving our students when we fail to show the shortcuts that we take. Of course, corners have to be cut because life is short and we cannot wait upon certainty. I am not defending compromises, but I am asking that we acknowledge that rigor is deeply connected to them in our own professional practice.

Rigor, Persuasion, and Rhetoric

I know that I am not alone in trying to push for inspecting aspects of rigor in qualitative research. Sandelowski (1993) for example, recognizing "the danger of succumbing to 'the illusion of technique'" (p. 1), argued that "rigor is less about adherence to the letter of rules and procedures than it is about fidelity to the spirit of qualitative work" (p. 2). True to a certain extent, but too ephemeral for me. I think rigor refers to more than the spirit of the research, whether qualitative or quantitative. As we have seen, an element of rhetoric seems to be lurking in some of the steps we take in our research. Some argue that the element of rhetoric in quantitative research is of significant proportions:

> The language of statistics is but one form of rhetoric; however, it is a rhetoric that, for certain audiences and in certain circumstances can be more compelling and more functional than a case study, poem, or autoethnographic report. (Gergen & Gergen, 2000, p. 1033)

The term "rhetoric" may have unjustly received bad press. Although the term is sometimes used to reflect a tone of insincerity or exaggeration, its origins are in the work of Isocrates; and its elaboration during the Renaissance by Erasmus and others (Shrag, 1982, p. 271) gave it its distinctive meaning of argument and persuasion. As Shrag puts it in his discussion of the traditions of knowledge:

> The rhetorical tradition realizes the limitations of philosophical argument as a vehicle for persuasion, especially when addressed to those who lack the training to follow the arcane, arid argumentation relished by that tradition. The rhetorical tradition recognizes a fundamental fact, namely, that people are creatures of flesh and blood, of passionate desire and aversion. (p. 272)

I have already made the point that research is about persuasion, and so is rhetoric. My concern is that we come clean about this and recognize rhetoric as part of our professional work. As Shrag notes, rhetoric is a tradition of knowledge that has been "the most influential tradition in European and American schools since the Renaissance" (p. 275). Once we have accepted that research is about persuasion, our task as researchers and graduate supervisors becomes one of acknowledging the place of rhetoric in discussions of the rigor of research, because our students need to know what is rhetoric and what is not, and they need to know what is poor rhetoric and what is good.

Envoi

I have tried to show that we need to replace talk about reliability and validity with a concept that recognizes that the value and purpose of research lies in human affairs. "Rigor" seems to do this, most especially when we understand that rigor has several facets. By promoting the idea of rigor and its facets, we might discourage students of research from reliance upon checklists about reliability and validity. Part of the danger of checklists is that they tend to sanitize research. The lists may remind us of the smaller pieces, but they contrive to teach the novice researcher that the enterprise is removed from human frailties. Research must not be "washed too much" in method texts, no more than we should treat such texts as biblical authority, far less as Decalogue. The facets of ethics, professionalism and rhetoric tell us plainly that rigor is very human.

To end, here is Auden's Hermetic Decalogue, the last 4 of the 29 stanzas of "Under Which Lyre":

Thou shalt not do as the dean pleases,
Thou shalt not write thy doctor's thesis
    On education,
Thou shalt not worship projects nor
Shalt thou or thine bow down before
    Administration.

Thou shalt not answer questionnaires
Or quizzes upon World-Affairs,
    Nor with compliance
Take any test. Thou shalt not sit
With statisticians nor commit
    A social science.

Thou shalt not be on friendly terms
With guys in advertising firms,
    Nor speak with such
As read the Bible for its prose,
Nor, above all, make love to those
    Who wash too much

Thou shalt not live within thy means
Nor on plain water and raw greens.
    If thou must choose
Between the chances, choose the odd;
Read *The New Yorker*, trust in God;
    And take short views.

References

American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostics techniques. *Psychological Bulletin, 51*, 2-7.

Auden, W. H. (1979). Under which lyre. In E. Mendelson (Ed.), *W. H. Auden: Selected poems* (pp. 178-183). New York: Vintage Books.

Aydelotte, W. (1971). *Quantification in history.* Reading, MA: Addison-Wesley.

Barry, C.A., Britten, N., Barber, N., Bradley, C., & Stevenson, F. (1999). Using reflexivity to optimize teamwork in qualitative research. *Qualitative Health Research, 9*(1), 26-44.

Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research in education: An introduction to theory and methods* (3rd ed.). Boston, MA: Allyn and Bacon.

Bruner, J, (1996). Narratives of science. In J. Bruner, *The culture of education* (pp. 115-129). Cambridge, MA: Harvard University Press.

Conant, J. (1957). *Harvard case histories in experimental science.* Cambridge, MA: Harvard University Press.

Davis, B., Sumara, D., & Luce-Kapler, R. (2000). *Engaging minds: Learning and teaching in a complex world.* Mahwah, NJ: Lawrence Erlbaum.

Dray, W. (1957). *Laws and explanation in history.* Oxford, England: Oxford University Press.

Fenstermacher, G. (1990). Some moral considerations on teaching as a profession. In J. Goodlad, R. Soder, K. Sorotnik (Eds.), *The moral dimensions of teaching* (pp. 130-151). San Francisco, CA: Jossey-Bass.

Firestone, W. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher, 16*(7), 16-21.

Gergen, M., & Gergen, K. (2000). Qualitative inquiry: Tensions and transformations. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed. pp. 1025-1046). Thousand Oaks, CA: Sage.

Goetz, J., & LeCompte, D. (1984). *Ethnography and qualitative design in educational research.* New York: Academic Press.

Goodman, P. (1960). *Growing up absurd: Problems of youth in the organized society.* New York : Vintage Books.

Gould, J., & Kolb, W. (1964). *A dictionary of the social sciences.* Glencoe, IL: The Free Press.

Hempel, C. (1968). Explanation in science and history. In P. Nidditch (Ed.), *The philosophy of science* (pp. 54-79). Oxford, England: Oxford University Press.

Heshusius, L. (1994). Freeing ourselves from objectivity: Managing subjectivity or turning toward a participatory mode of consciousness. *Educational Researcher, 23*(3), 15-22.

Hoepfl, M. (2000). *Choosing qualitative research: A primer for technology education researchers.* <http://www.curriculum.edu.au/tech/articles/choose.htm> (February 19, 2001).

Holt, J. (1964). *How children fail.* New York: Pitman.

Jackson, P. (1968). *Life in classrooms.* New York: Holt, Rinehart & Winston.

Janesick, V. (1994). The dance of qualitative research design: Metaphor, methodolatry, and meaning. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 209-219). Newbury Park, CA: Sage.

Kohl, H. (1967). *36 children.* New York: New American Library.

LeCompte, M., & Schensul, J. (1999). *Analyzing and interpreting ethnographic data.* Walnut Creek, CA: Altamira Press.

Margenau, H. (1950). *The nature of physical reality: A philosophy of modern physics.* New York: McGraw Hill.

Margenau, H. (1972). The method of science and the meaning of reality. In H. Margenau, (Ed.), *Integrative principles of modern thought* (pp. 3-43). New York: Gordon and Breach.

McMillan, J., & Schumacher, S. (1997). *Research in education: A conceptual introduction* (4th ed.). New York: Longman.

McMillan, J., & Schumacher, S. (2000). *Research in education: A conceptual introduction* (5th ed.). New York: Addison Wesley Longman.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Mischler, E. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies, *Harvard Educational Review, 60*, 415-442.

Moschkovich, J., & Brenner, M. (2000). Integrating a naturalistic paradigm into research on mathematics and science cognition and learning. In A. Kelly, & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 457-486). Mahwah, NJ: Lawrence Erlbaum.

Munby, H. (1983). A perspective for analyzing the significance of qualitative research: A response to Richard Heyman. *Curriculum Inquiry, 13,* 423-427.

Munby, H., Chin, P., & Hutchinson, N. L. (2000, April). *Co-operative education, the curriculum, and "working knowledge."* Paper presented at the Internationalization of Curriculum Studies Conference, Louisiana State University, Baton Rouge, LA.

Munby, H., Russell, T., & Martin, A. (in press). Teachers' knowledge and how it develops. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.): Washington, DC: American Educational Research Association.

Roberts, D. (1982). The place of qualitative research in science education. *Journal of Research in Science Teaching, 19*, 277-292.

Sandelowski, M. (1993). Rigor or rigor mortis: The problem of rigor in qualitative research revisited. *Advances in Nursing Science*, 16(2), 1-8.

Shrag, F. (1992). Conceptions of knowledge. In P. Jackson (Ed.), *Handbook of research on curriculum* (pp. 268-301). New York: Macmillan.

Simpson, J. A., & Weiner, E. S. C. (Eds.). (1989). The Oxford English dictionary (2nd ed,. Vols 1-20). Oxford, UK: Clarendon Press.

Soder, R. (1990). The rhetoric of teacher professionalism. In J. Goodlad, R. Soder, K. Sorotnik (Eds.), *The moral dimensions of teaching* (pp. 35-86). San Francisco, CA: Jossey-Bass.

Webb, R. B., & Glesne, C. (1992). Teaching qualitative research. In M. LeCompte, W. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 775-776). New York: Academic Press.

Winchester, S. (1998). *The professor and the madman.* New York: Harper.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

TM032879

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: EDUCATIONAL RESEARCH AS DISCIPLINED INQUIRY: EXAMINING THE FACETS OF RIGOUR IN OUR WORK[1]

Author(s): *Hugh Munby*

Corporate Source: Faculty of Education, Queen's University, Kingston, Ontario K7L 3N6

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature:

Printed Name/Position/Title: *Hugh Munby*

Organization/Address: Faculty of Education, Queen's University, Kingston, Ontario K7L 3N6

Telephone:

FAX:

E-Mail Address: munbyh@educ.queensu.ca

Date: April 12, 2001

(over)