INCORPORATING STAKEHOLDERS IN
STANDARD SETTING: WHAT'S AT STAKE?

Darryl M. Hunter
Planning and Evaluation Branch,
Saskatchewan Department of Education
Regina, Saskatchewan

Trevor J. Gambell
Department of Curriculum Studies, College of Education,
University of Saskatchewan
Saskatoon, Saskatchewan

Abstract: In this article, we discuss the roles played by participants and
the processes involved in a provincial educational standards-setting exercise for a large-scale assessment of student skills.
The perspectives offered are those of a policy maker outside the
exercise and of an insider participant who describes his experience of the standard-setting exercise from within. A number of issues involved in selecting delegates, in empanelling stakeholder representatives, and in designing standard-setting exercises are considered. Panelists are characterized as anonymous jurors rather than as subject-matter experts. An exercise will reconcile competing organizational and social values when defining standards as points for educational decision making. We conclude by describing alternate notions of representativeness, stakes, and significance.

Résumé: Cet article discute des rôles joués par les participants et des
processus mis en œuvre lors d’un exercice de formulation de normes en éducation dans le cadre d’une évaluation à grande échelle des performances des élèves. Deux perspectives sont adoptées, celle d’un analyste de politique ne participant pas directement à l’exercice et celle d’un participant décrivant l’expérience de l’intérieur. Plusieurs questions sont abordées dont le choix des participants, l’inclusion des principaux intervenants, et l’organisation même de tels exercices. Les participants sont conçus plutôt comme des jurés anonymes que des experts dans leur domaine d’études. Un des buts de ces exercices est de réconcilier les valeurs organisationnelles et sociales en établissant des normes comme points de convergence lors de prises de décisions en éducation. L’article se termine en suggérant de nouvelles définitions des notions de représentativité, d’enjeux, et de signification.
Our article explores issues of stakeholder involvement in evaluation, described from two different vantage points in a standard-setting process. Darryl Hunter offers a policy and planning perspective as Director of Assessment and Evaluation for the Saskatchewan Department of Education which sponsored, planned, and conducted the Saskatchewan 1998 listening and speaking assessment (Saskatchewan Education, 1999a) and conducted the national scoring for the reading part of the 1998 SAIP literacy assessment (Council of Ministers of Education, Canada, 1999). Darryl was not a participant in the standard-setting session, but was pivotal in its design. Trevor Gambell provides the perspective of a participant stakeholder, delegated as a literacy scholar from the College of Education at the University of Saskatchewan, who was invited to participate along with 25 other panelists.

Darryl's and Trevor's dialogue below stems from their experience in standard-setting processes associated with both provincial and national assessments (Hunter & Gambell, 1996). This evaluative procedure has been of interest to both of us in a scholarly and professional sense, because criterion-referenced standard setting for large-scale assessments explicitly garners the exogenous judgments of panelists rather than drawing on the endogenous comparisons of typical student performances in norm-referenced testing procedures. Our exchange centres on a three-day provincial standard-setting exercise that determined both opportunity-to-learn standards (desirable learning processes and conditions) and performance or test outcome standards (desirable products of learning) for a provincial listening and speaking assessment carried out in 1998. At the same session, stakeholder representatives also set provincial opportunity-to-learn standards for reading and writing using questionnaire data from the national School Achievement Indicators Program (SAIP) literacy assessment conducted in 1998 by the Council of Ministers of Education, Canada (Council of Ministers of Education, Canada, 1999; Saskatchewan Education, 1999b).

Standard-setting exercises in Saskatchewan and on the national level have, during the past five years, anonymously empanelled both experts and non-experts drawn from a variety of stakeholder organizations to define public expectations for student performance. At these exercises, panelists are oriented in the premises and procedures of the assessment, provided with a wide array of test evidence, engaged in multiple-round voting and group discussion procedures using Delphi-like techniques, and charged with independently pre-
scribing desirable percentages of the student population who would or should achieve criterion levels of performance. The resultant standard, conceived as a point of educational decision making and as a comparator for provincial and national indicators of systemic outcomes, is an important instrument of public accountability.

STAKES AND STAKEHOLDERS

Darryl: Civil servant employees of the Saskatchewan Department of Education have, as a basic job requirement, the mediation and accommodation of diverse interests in education. In fact, it is an explicit function of the department, enshrined in our mission statement. Therefore, as a public servant, I take as a central presumption that all evaluative activity in the province should draw in some way on sometimes divergent stakeholder values. Yet I am often perplexed that the notion of a “stakeholder” remains a rather wooly concept in evaluation, particularly for the purposes of public accountability. A variety of collateral terms colour the literature on stakeholder-based approaches to evaluative decision and policy making, ranging from Lindblom’s (1959) “proximate policy-makers” through to Patton’s (1997) “primary intended users” of evaluative information. Many of the theoretical and operational dimensions of stakeholder participation have not been extensively explored, such as stakeholder definition and selection, the nature of meaningful participation, and the role expectations of participatory evaluators (Greene, 1988).

It was Weiss (1972) who first emphasized the diverse interests of stakeholders in an evaluation, whom she subsequently defined in all-encompassing terms as “members of groups that are palpably affected by the program and who therefore will conceivably be affected by evaluative conclusions about the program, or the members of groups that make decisions about the future of the program.” (Weiss, 1983, p. 84). That’s not a very helpful definition for evaluations conducted for public accountability purposes, since society as a whole may be “palpably affected.” For Robert Stake, the term “stakeholder” was coined to help the evaluator answer the question, “Whose side you are on?” (1981, p. 38) and includes those “who are running the program, those participating in the program, those patrons of the program” (1975, p. 36). He, of course, would be skeptical that large-scale assessments in the quantitative mode can be sensitive to local stakeholder interests. For Scriven, a stakeholder is one among many competing value systems that might be considered,
but above which the evaluator must rise to make summative and presumptive judgments in the name of the public interest. Indeed, he has gone so far as to describe the derivation of values from interest groups as “the ‘lackey’ view of evaluative research” (1986, p. 23).

*Trevor:* The thought or accusation of being a “lackey” in the standard-setting process in Saskatchewan — in which I participated as a panelist — would have been anathema to the people gathered together for the September 1998 exercise. There was a practical purpose in this exercise, not a political one, and each of us considered ourselves enabled in the sense that we were designates of a stakeholder organization but, at the same time, independent agents free to debate the issues that arose from our individual value perspectives. We felt empowered to make educational decisions that would legitimate education policy and evaluation procedures for Saskatchewan.

The standards panel comprised 26 people, not including session leaders and technical personnel from Saskatchewan Education. Fourteen members were practising teachers, and the other twelve were educational stakeholders and industry representatives. We were expected “to take a common-sense and collaborative approach to setting the standards,” in the words of the facilitator who animated the sessions. The grade levels we dealt with were grades 5, 8, and 11 provincially, and grades 8 and 11 nationally. For the provincial portion (listening and speaking), we were assigned to one of the three grade-level panels, along with one or two session leaders and a computer associate. For the national portion (reading and writing), the grade 5 panel was divided and reassigned to the existing grade 8 and 11 panels. Setting outcome standards was described by the facilitators as a “medium-to-high-stakes” activity. Because students’ performances on the listening and speaking assessment were anonymous and would not contribute to their school marks, these were considered “low-stakes” assessments for students. Students’ only motivation to do their best was “personal pride for the common good.” The same can be said for the national reading and writing assessment.

*Darrryl:* Actually, the underlying question of motivation for participation applies to our view of what’s at stake. The notion of “stakes” is key because how it is defined will in turn suggest the composition of a standards panel. Stakes can be conceived alternatively as: the consequences from a decision in terms of career or life prospects; a
personal or organizational investment in a decision; a legislatively assigned responsibility for action devolving from a decision; the risk associated with political or media exposure; metaprofessional or metaoccupational pride; or simply as the desire to maximize potential for youth or the product of one’s program. The stakes, in short, vary according to the holder. In fact, many questions remain about who should set standards. Is an educational standard a professional responsibility, a social construction, or a bureaucratic creation? Out of undoubted concern that bureaucrats can adequately reflect the full range of interests in society, many scholars (Hambleton & Powell, 1983; Jaeger, 1989; Shepard, 1980) suggest that standard setters be drawn from different constituencies, so that the standard-setting panel can systematically incorporate different values and areas of interest.

EXPERTS AND EXPERIENCE

Few guidelines have been formulated for selecting these panelists. At issue is the degree to which panelists should have expertise in the subject matter being tested, experience in the curriculum design and instructional policies that prevail in schools, knowledge of the attributes of the population being tested, or an understanding of the maturational characteristics of youth. Likewise, we do not know whether a panel of classroom teachers will produce more appropriate standards than a mixed panel of educators and non-educators. Some suggest a standard produced by a blue-ribbon panel, like the French Académie, may be more credible than that produced by an anonymous jury. Others argue that the panel should consist of those who have a stake in the decisions that result from the standard that is defined and not only those who understand student capabilities. As a university professor, you certainly have a stake in student competence, but did you think that your status enhanced your stature in the committee?

Trevor: We were never led to believe that we were a blue-ribbon panel; our presence was as delegates of the stakeholder groups identified by Saskatchewan Education. When we met as a combined committee, we did not introduce our affiliations; in grade-level panels, we did. Most of us did not know each other, and it was reassuring to be anonymous. That anonymity permitted us to speak freely and personally, from our own experience and not as an “official” institutional voice or perspective, whatever that might be. In the case of teachers and myself, it would be difficult to formulate an institu-
tional position. At a university, allegiance is likely to be to a professional, subject-matter philosophy rather than a mission statement. I assumed we’d been selected for our educational experiences and curriculum and instruction expertise rather than as a designated stakeholder mouthpiece.

Panelists in my group certainly took their task seriously and drew on their experience with young people as adults, as businesspeople, as educators. Members understood the importance and implications of the standards they were setting — that these expectations would have the potential for significant impact on student assessment and provincial educational achievement levels. Teachers outnumbered other members, and although teachers did not speak or vote as a block, they individually complemented each other’s perspectives. I would say that the teacher perspective subtly dominated the parent one. Parents, of course, want the best for their children, but teachers tended to bring a level of reality to the discussion, based on their accumulated experience in heterogeneous classrooms.

Darryl: As you imply, I too ask, who is an expert in prescribing desirable performance for youth? Jaeger (1991) draws an overly easy dichotomy between standards established by policy makers whose authority derives from position rather than qualification and standards based on specialist judgments concerning the difficulty of test items. For the latter type, Jaeger asserts, it is necessary to recruit panelists who have expertise in the domains being tested and in the roles sought by successful examinees. Certainly, in “high-stakes” assessments, potential court cases about equality of treatment may be a consideration for the policy maker, and indeed have been in the United States. There, expert viewpoints on student prowess have proven persuasive in the legal system (Jaeger, 1989). Yet a public accountability assessment invites public and not just expert involvement in standard setting. With random-sample assessments, it is not the law courts but the court of public opinion that is interested in the standard. The tests may be “low-stakes” for students, teachers and schools, but are “high-stakes” for ministers of education and senior public administrators responsible for maintaining education quality. In fact, the question of who should adjudicate the outcomes from an entire education system remains contentious but critical, since standards are part of sustaining confidence in public education. So, with regard to stakes, did conflicting interests make you feel as if burnt at the stake, or did the various stakeholders develop a *modus vivendi*?
Trevor: At first meeting, panelists seemed to be a motley crew. The one common ground seemed to be both professional and/or personal interest primarily in literacy, and overall in education. The involvement of stakeholder organizations in policy development and planning for large-scale assessments in Saskatchewan may well be unique to the province. Saskatchewan enjoys longstanding co-operation and collaboration among its educational stakeholders which include the provincial department of education (Saskatchewan Education), the Saskatchewan Teachers’ Federation (STF), the Saskatchewan School Trustees Association (SSTA), the League of Educational Administrators, Directors and Superintendents (LEADS), and the two provincial universities through their faculties of education. When curriculum advisory committees and assessment steering and development committees are established, invitations customarily go out to these agencies to appoint representatives. Teachers usually constitute the greatest number of participants. For the standards-setting panel, this same selection policy was followed, with some important additions. Such a broad mix brought multiple and not always compatible perspectives to the committee, which were evident on several occasions during the three days.

VOICE OR VOTE

Darryl: Actually, we spent quite a bit of time debating, both within the department and with another senior policy advisory committee of stakeholders, the relative weighting of committee membership. One question was whether all panelists should have a vote, as a decision-making right, or whether a distinction should be drawn with those who would have only a voice, as an expression of influence. What should be the combination of voice and vote? The policy problem arose when we set provincial mathematics standards in 1995 and 1997 with both mathematics specialists and non-specialists; large gaps were evident in virtually every dimension of mathematics between results and the criterion-standard that was set. Senior government officials and executive members of stakeholder organizations did not believe that the standards were credible or realistic because of these consistent gaps. Some felt that panelists had “built castles in the sky,” and others questioned the judgments of non-educators. I wondered whether the exercise had been misconceived because of the question to which panelists responded, and worried that the standard — far from building public confidence in schools — would undermine our concerted attempts to improve the education system. Even though subsequent national assessments have
revealed similar normative shortcomings in the mathematics achievement of Saskatchewan students, and even though language arts standards have revealed both strengths and weaknesses, questions have remained in policy circles about the trustworthiness of criterion-referenced standards.

Accordingly, for the 1998 exercise, we decided at a policy level to accord a vote to all stakeholder delegates for the opportunity-to-learn standards but agreed that only educators would have the final vote for the test outcome standards, with non-professional voices heard throughout the process. Yet we wondered the whole time about the consequences of disenfranchising non-professionals in the standard-setting session, and for only a crucial part of the session at that. How did this work in practice?

Trevor: It became an issue during the culminating stages of the exercise and threatened the cohesiveness established up to that point among the panelists. The panel was informed at the end of round two, that the subsequent round three would be voted on only by the teachers in the panel, those officially representing the provincial teachers’ association. This took our panel members by surprise, because it was announced only during that morning’s plenary session. The explanation provided was that this agreement had been made by the senior advisory committee that included representatives from all of the stakeholder groups. The teachers in our panel were astounded and embarrassed by the unexpected announcement. In effect, the teachers set the standard in the round three voting in each instance, although all panel members participated and voted in rounds one and two.

Therefore, I became interested in seeing if and how the vote differed between rounds two and three. Those who had voted high in round two were concerned that they could not adjust their provisional standards in another vote and, if they felt strongly about their vote, that they would have to skillfully convince the teachers on the panel of the need to adjust upwards their percentage vote on round three. But the actual percentage differences between rounds two and three were slight, no more than 5% in the initial dimensions adjudicated and much less than that for the final skills dimensions. Thus I am sure that the teachers voted as consistently as the panel in round two in each instance, except where our whole panel discussion convinced them to make adjustments. In other words, this potentially divisive procedural rule did not turn out to have distorted the panel’s discussion nor its overall expectations.
Darryl: I should perhaps offer background. When prescribing the overall standards committee composition, we believed that teachers should have the preponderate weighting, both because they will be the primary users of standards at the school level and because they were seen as best positioned to ascertain the abilities of youth. Nevertheless, school boards have a key role in allocating resources, so trustees needed to have a substantial block of votes. We tried to strike a balance by including singular representatives from the other agencies to ensure those perspectives were brought to the table. However, parental groups insisted that they should have a larger role than in the original plan, so we accorded additional seats at the table. Actually, school administrators were concerned after the exercise that their influence in school affairs was not substantially mirrored in the panel composition, particularly in comparison with parents and trustees. So the weighting of stakeholder delegates was brokered at a policy level — not so much according to questions of power but more in terms of ensuring that the overall picture mirrored holistically the array of interests in educational policy making.

Nor should we forget the underlying motivation for involving a variety of publics in setting a standard, which both incarnates and refracts a variety of values. It is not only to garner a broad array of social values, but also to maximize the impact of a standard. I have been strongly influenced by a poll conducted in Saskatchewan in the early 1990s that asked the general public who they deemed most credible as purveyors of information about the quality of education. Most people trusted teachers the most. Next in terms of credibility were school administrators, such as principals; the general public trusted school board officials to a lesser degree, but more so than civil servants such as Department of Education officials. Those deemed least credible, as porte-paroles of educational quality, were politicians. In other words, the public placed greater confidence in those educational authorities closest to the classroom setting, not those toward the hierarchical apex. This realpolitik therefore suggests that we might best address issues of public accountability by giving classroom teachers the tools, such as standards, to communicate with parents and property tax interests.

CONFLICT AND COLLABORATION

Trevor: The rationale that you’ve just provided would have been, I believe, acceptable to the panelists. I happen to agree that professional educators, who deal with curriculum and students on a daily
basis, are best able to interpret curriculum objectives and test outcome results; thus they have a high stake in the levels set for test performance outcomes in terms of curricular and instructional impact. Teachers in particular were distrustful of the business perspective; parents were noncommittal and most likely to defer to the position taken by teachers. There is no doubt in my mind that the parents among the panelists put greater faith in teachers than did other non-educator panelists. Business and industry have been outspoken critics of public education, and their motives are greeted with skepticism by many educators. A predominant voice and/or vote from the business and industry sector would not be advisable in this sort of exercise.

Darryl: That’s interesting, because these same tensions have been manifest in other provincial policy arenas during the past five years. Standard setting in Saskatchewan originates in the work of a Minister’s Advisory Committee that reviewed high school education in the province in 1994. Significant dissent was expressed within the committee when it made recommendations relating to standards and testing. A business representative called for universal testing to extend beyond the Grade 12 level, while Aboriginal committee members opposed the use of standardized, paper-and-pencil tests as incongruent with the diverse school situations in the province (Saskatchewan Education, Training and Employment, 1994).

A subsequent 1996 symposium amplified these conflicting perspectives and showcased the kaleidoscope of opinion in the province about the standards issue (Saskatchewan School Trustees Association, 1996). An official of the Canadian Federation of Independent Business asserted that “educational standards are important to Saskatchewan business to ensure that minimum competencies, understandings and skills are consistently assured by graduates of our school system as part of a quality labour force” (p. 5). He was admittedly blunt in reporting that “business people do not want to do some of the ‘most basic product recall work’ on behalf of our educational factories” (p. 5). Likewise, a trustee speaking on behalf of the Saskatchewan School Trustees Association advised those in attendance that “we must agree on accountability measures that will tell us how well students are meeting objectives. If we don’t develop such measures, outside pressures will force them upon us [...] Standards will help us answer the question, ‘How do we know we are doing a good job?’” (p. 65). Similarly, a Saskatchewan Department of Education official asserted that “it is virtually impossible to argue
we shouldn’t have standards” (p. 42), but emphasized that opportunity to learn was perhaps more important than focusing on outcomes.

However, many doubts were expressed in that 1996 forum. Speaking on behalf of the League of Educational Administrators, Directors and Superintendents, one administrator cautioned that the province must remain “loyal to standards development that stresses the processes of learning” as opposed to only product skills (p. 63). For a Saskatchewan Teachers’ Federation representative at the forum, the call for standards was misguided. Likening teacher-student bonds to a farmer’s attachment to the land, she asserted that, “The relationship between teachers and students, at its best, is a marvellous, even sacred thing. It cannot be captured in lists of outcomes, in scope-and-sequence charts, in taxonomies of standards, in rubrics” (p. 60). So, Trevor, you’re saying that these same tensions were manifest within your panel?

Trevor: Yes, perhaps because this was the first, and only, time that this committee would convene for this purpose. Because members knew the task was to be accomplished in three intensive days, common understandings and working relationships had to be established at the outset. A major concern was the comfort of individual members and confidence in their own qualifications to set standards for student outcomes and opportunities. I sensed that the greatest concern came from parents, who possibly felt somewhat intimidated by the predominance of professionals in the panel (Fitzpatrick, 1989). The facilitators and other participants quickly assured the parent representatives that their parental experience was respected and considered as equally valid as professional experience; there was a strong though implicit desire to develop a respectful and tolerant atmosphere in the group.

Darryl: Could I interject here, Trevor? Were you not, as a panelist, simultaneously an academic, a stakeholder representative, and a parent? Don’t panelists wear multiple hats at a standard-setting session, often operating simultaneously as parents, trustees, and small business owners? Likewise, teachers will come to a standard-setting session simultaneously as professionals with decades of experience with youth, as parents, as perhaps a home business owner, and even as a school administrator — because they combine those roles in their lives. In other words, I am suggesting that stakeholders are not homogenous collections of two-dimensional figures whose positions can easily be labeled. Organizational boundaries, more-
Individual standard setters usually hold different perspectives on what a standard signifies; it functions as a sort of Rorschach blot in education. Standard setting can thus be viewed in many ways — as a forum where the match between what educators intend to do and what they actually have done is deliberated; as the nexus where educational ideals meet reality; as an exercise which gives voice to the underlying appreciative knowledge of professionals and non-professionals about the potential for youth; or as a public policy arena where competing stakeholder interests and values are drawn together and reconciled to define a decision-making point for public accountability purposes. From my perch, it is largely an exercise in clarifying and reconciling individual, group, and social values. My observation is that standard setters must not only reconcile those different perspectives in an intra-group sense, but also internally and individually during a standard-setting session. We are back to the old question of principal and agent. Did you believe that you were only carrying the brief of your dean of education or the university board of governors to the standards table?

Trevor: There is no doubt that panelists wore more than one hat during their deliberations. Teachers sometimes doffed their educator hat for their parent hat; parents had only one hat to wear. A concern of the group was the inclusion of a business and industry representative, who did not show up for the first day, but who was there for the second day (but not the third). Professional educators considered the business and industry sector’s public statements, about lack of student preparedness and skills development to take their place in post-secondary employment, as a personal attack on their professional abilities. They felt that business and industry do not understand the nature of today’s students (diverse ethnic, racial, religious, and socioeconomic background), curriculum developments (gender inclusion, process curricula, alternative assessment) or teaching methodologies (group work, collaboration, co-operative learning). As well, some educators believe that business and industry is shirking responsibility for on-the-job training and apprenticeship, expecting schools to prepare students at public expense to undertake any number of diverse employment roles at the end of the high-school years. Although these perspectives remained tacit, they undoubtedly were behind the discussions that arose when the business representative, who characterized himself as an adult edu-
WOULD OR SHOULD

Darryl: I want to ask about another contentious policy issue on the nature of the standard. Should a standard prescribe a short-term target, an ideal, or a realistic estimate in terms of the current range of student skill or ability? More pointedly, isn’t the nature of the standard determined by the question that standard setters answer? In American judgmental procedures, panelists are asked to estimate the percentages of students who “would” answer a test question correctly, as in Angoff’s (1971) method, or the percentages of students who “should” answer a question correctly, as in Jaeger’s (1982) method. A “would” question may produce a realistic standard, whereas a “should” question asks for a formulation of student potential in optimal circumstances. Rather than anticipating performance, a “should” question may ask panelists to provide aspirations rather than estimates. As a target to aim for, the “should” standard may become unattainable (Hunter, Randhawa, & Owen, 1996). Groucho Marx stated this problem succinctly when he quipped, “I have my standards, and some day I hope to live up to them.” How did your committee deal with this issue, Trevor?

Trevor: The issue of standards “rooted in reality” was a by-product of the two perspectives given panelists as ways to derive standards. We were provided with written information that described these two perspectives. The “should” perspective derives standards based on desire — on our hopes for the best for each student and on our thinking that each student is capable of reaching the highest possible performance level. Such hope is projected onto the whole population of students. This perspective considers possible societal needs for the future, with our aim being to promote improvements in student learning and performance. The “would” perspective derives standards based on expected or anticipated outcomes. This perspective is characterized by consideration of the full context of the assessment and weighing all the evidence in the assessment (not just test returns). The facilitators were asking for information on performance and opportunities to guide improvements in the education system.

Those standard-setting sessions for student opportunities and outcomes relating to listening and speaking sought to establish desired results, the “should” perspective, after careful consideration of the
“would” perspective. The goal was to arrive at desired opportunity and outcome standards grounded in the complexities and realities of learning. Our written guidelines cautioned us against setting idealistic standards or standards of hope. But given our panel’s consideration of existing and new English language arts curricula — with the latter’s equal attention to listening, speaking, reading, and writing — we were led recurrently toward the standards-of-hope perspective as we anticipated the impact of a new curriculum, with which a number of us had experience at development or pilot stages.

Another concern that arose in the early stages of the exercise was the curriculum reference to be used in our deliberations. A new curriculum just being piloted at the grade 11 level has listening and speaking as a process domain equal in emphasis to literacy (reading, writing, language study). But the existing curriculum for grade 11 stresses literacy and literature, and the educators in the group were concerned that students would have had little opportunity to develop listening and speaking skills. This discussion extended to interpretation of our roles. Did we see our function in standard setting as signaling the current state in light of the present curriculum (thus not necessarily lack of student skill but lack of opportunity) or as a beacon for future change in light of the new, incoming curriculum? We concluded that the data under consideration derived from a learning assessment, not from a curriculum evaluation. Thus we felt assured that the standards we were about to set would be independent of existent or incoming curriculum. Toward the end of the exercise, our facilitators advised us to think about what we should expect students to be able to accomplish in the next two years. Therefore, a relatively high standard set now does not necessarily signal a deficient skill level in light of existing curriculum.

_Darryl:_ These issues are important. Unless a stakeholder panel grounds its views within the realm of possibility, then its considered judgments will be disregarded or dismissed by officials who must dwell in the realm of feasibility and practicality ... and accountability. Measurements need some form of comparator to make them meaningful; standard setting is meaning making with data by members of society and hence constructivist in its dynamic. Educational standard setting with test results might thus be seen as a form of sociometrics. Which brings me to another question, “How trustworthy do you think your committee’s work was?” Should a minister of education or a school administrator or John Q. Public have confidence in the standard as being credibly and fairly set?
CONSENSUS AND CONSISTENCY

Trevor: Definitely, the prevailing ethos in the standards panel was to reach a common understanding, but not necessarily a consensus. There was room for divergence but, when all numerical votes were gathered, there was little variance among individual percentages in the final rounds. In earlier rounds, the variance was initially quite wide but narrowed as discussion followed each round of voting and another vote was taken.

Our panel wrestled with the definitional problem of “rooted in reality.” Clearly the opportunity-to-learn standards created greater divergence and discussion than did the performance outcomes. Because of the highly subjective nature of opportunity and the interpretive difficulty of determining how reading or writing were understood by student respondents, our panel needed to discuss opportunity-to-learn measures more fully when setting expectations. Our facilitator admonished us repeatedly to consider expectations that were “rooted in reality,” which meant consideration of students working part-time while attending school full-time, a variety of alternative lifestyles, sports, and other activities taking precedence over homework, and students repeating grade 12 to improve their academic grades. Our panel repeatedly returned to the “rooted in reality” construct as we sought to balance our expectations with well-intentioned idealism.

Darryl: Are you saying, Trevor, that it is much easier to achieve group consensus when panelists consider student performance or opportunity in the ideal? Is that not groupthink? — that a group becomes so engrossed in its ideal that members fail to consider a variety of other, pragmatic alternatives?

Trevor: I do not think that groupthink is manifest in co-operative deliberation, or even in consensus. I believe that in our panel, consensus was neither sought nor accepted as a goal that the panel should attain or even strive valiantly toward. I think that the individuals in our panel were satisfied in voicing their perspectives and in exercising their votes after satisfactory discussion. They had to feel comfortable that the final panel decision was one wherein their individual perspective was maintained, and that they were able to return to their workplaces, homes, and constituencies and defend the decision of the panel without being themselves defensive.
Given the important distinction between the “should” and “would” perspectives within standards setting, and the recurrent need within our panel to temper individual idealism when settling on desired outcome standards, one asks whether teachers as practising professionals are any more or less idealistic in their expectations than other stakeholders. In our committee, discussion kept coming back to the optimism for student learning promised by the new English language arts curriculum and complementary inservice education for teachers. This optimism bred idealism; teachers in our group expected that in the future students would and should do much better in listening and speaking and literacy assessments because they will then have been exposed to new curriculum and reformed instruction. Discussion evolved in the direction of setting a standard that was reasonably achievable at this time with the existing curriculum but which would also leave room for heightened attainment as the new curriculum permeated schools.

But despite the positive anticipation of a new curriculum, teachers’ voting was not, as a group, very much different from that of other panelists. Convergence in voting from the first to the second and subsequent rounds was a regular feature of how the balloting unfolded. I believe that all adults who work with young people will be idealistic in their hopes and expectations for the next generation. This, it seems, is part of the human condition, to hold higher hopes for the next generation. Teachers, many of them parents as well, blend their professional expectations and their parental ones, possibly with a heightened anticipation for all students.

_Darryl:_ I still wonder whether panelists’ expectations might balloon irrationally in the pursuit of consensus, since committees often reach agreement by seeking more abstract or lofty ground (Fitzpatrick, 1989). The standard-setting procedures we have derive from American methods devised for minimal competency testing. All aim for agreement among panelists, because psychometricians believe that consistency denotes inter-rater reliability, which in turn connotes trustworthiness.

Yet consistency is not the same thing as consensus. There may be degrees of engagement within a consensual decision, ranging from apathy to acquiescence to consent to consensus to commitment. We have all sat on committees where peoples’ enthusiasm for a decision varies dramatically. Group pressures can seduce or coerce an individual within a range of values that has been called “a zone of indif-
ference.” In other words, there is an affective element that may mean there is a meeting of minds about a decision, but not a wedding of wills. As such, consensus means not only dissolving contradictory views on acceptable student performance, but also extinguishing individual positions and fostering group resolve.

Trevor: No, I am not saying that consensus means extinguishing individual positions and fostering group resolve. Such an abdication of personal responsibility could be viewed as playing the chameleon, or worse, of being a pawn of either government or institutional policy. Rather, individual panelists must be able to retain personal integrity, through maintaining their individual position, yet feel committed to the group process. That is why consensus may just be the wrong expectation for this type of group dynamics which, in our panel, evolved into a tacit process of professional construction of standards rather than a social or other (personal, representative stakeholder) standard.

Darryl: I am more ambivalent about the necessity for reliability as a precept in standard-setting design (Moss, 1994). Achieving consensus or making substantial progress toward it, Cizek (1996) notes, may be an inappropriate way of determining whether a standard-setting session has been effective. If participants have been selected to reflect a plurality of values, iterative standard-setting procedures may reveal little or no movement toward consensus. Failure to achieve consensus may actually be evidence of trustworthiness, in that “intended diversity was retained throughout the decision-making process” (p. 20).

On the other hand, if consistency in outlook is unnecessary, it might be asked if a standard-setting panel need be convened at all. Opinion makers could be surveyed at home to garner their pre-judgments without the tempering or illuminating influence of assessment information and exposure to the reasoned opinions of others in a standards exercise. Likewise, a collaborative and consensual decision is desirable from a public policy perspective. If the standard is to be used as a lever for making changes in the educational system or classroom practice, a meeting of minds among participants may foster a concerted commitment to act on the results to effect educational improvement. In fact, a mathematically derived consensus has been deemed unsatisfactory by participants in previous Saskatchewan standard-setting exercises; they felt a need for an actual meeting of all minds involved in the session (Jones & Hunter, 1996).
PROFESSIONAL OR SOCIAL VALUES

*Trevor:* The composition of the 26-member panel raises the question of whether the standards are a social or professional construction. Saskatchewan Education itself seems not to have a position on this issue, with its tacit agreement that the products of public schooling are workplace employees, while at the same time agreeing that only teachers will set performance or outcome standards for language arts. What is your reading, Darryl?

*Darryl:* I doubt whether we can definitively say that standards are either professionally or socially constructed, because I do not think that panelists, even though appointed by stakeholder organizations, operate exclusively as agents of the interest group, or of society, or of a profession. I think they make individual, deliberative decisions that involve the complex reconciliation of a variety of different values at different levels — personal, group, organizational, subcultural, and societal. On one hand, panelists must make personal decisions that may or may not be consonant with those of their sending organizations (Liedtka, 1989). On the other hand, they must also reconcile their organization’s values with those of the subculture or society of which they are a member (Hodgkinson, 1991).

*Trevor:* All members of society have expectations of public schools and of schooling. Who has a voice in determining standards that students should be expected to meet is a clear indication of the department of education’s perspective on standard setting and hence the purpose of public education. To produce citizens who are able to contribute and take on leadership and participant roles in a civil society is one purpose. The other is to produce graduates who are able to fill the needs of the business sector which, in terms of social economic policy, is not often sympathetic to liberal socialist policy.

As I said before, when one asks who votes in the standards-setting process, it seems clear that professional educators, especially teachers, have the dominant voice. This perception was reinforced by giving teachers the exclusive vote in the third and climactic round of setting outcome standards. My conclusion is that standard setting in this exercise was predominately a professional construction. I wonder how the process may have unfolded had recent graduates and/or employees representing a range of professions and trades been involved.
Darryl: Another way of exploring the issue of whose values predominate in a standard-setting session is to ask about the bases upon which panelists make their decisions. Where do panelists derive their estimates of student performance? Some scholars are concerned that “they seem to be pulled out of thin air.” How do panelists arrive at their judgments?

Trevor: I believe that most panelists arrived at their judgments through experience in working with a variety of young people over time — that is, developmentally — and through informed opinion and belief. This decision-making rubric informed both parents and teachers on the panel. However, the person from the business sector did not, I now think, have the necessary experiential background, either professionally or personally. So he had to default to a generalized and decontextualized response which came from a critical rather than conceptual base.

This perspective suggests that individuals informed similar to teachers would be effective panelists. Social workers, recreation workers, sports coaches, and others who work with young people over time might also be considered as panelists. Yet the potential to effect educational change would not be there without professional educators being involved.

Darryl: I am not sure we should automatically exclude those without the conceptual base; for example, those unversed in the law often render judgments in the courtroom as jurors. In my experience, standard setters in Canada may better be described as “jurors” rather than as “judges,” a term which is used in American psychometric literature. The label of “judge” suggests specialized expertise and advanced professional preparation in an academic discipline, whereas public administrative standard setting in Canada draws on more general, lay qualities of common sense, ability to approach a wide variety of evidence in an unbiased manner, and good judgment sought in a typical courtroom juror. A better comparison may be with the jury of 12 people drawn from a variety of walks of life, and without legal training, used in the legal system. If impartial non-experts are deemed acceptable for making “high-stakes rulings” in criminal and civil actions, then a panel of non-experts should analogously be sufficient to represent the informed, “low-stakes judgments” of citizens as part of a learning assessment.

Of course, there are differences between a jury in the legal system and a panel in defining educational standards. For one, standard-
setting panels work under administrative law to consider a wide variety of evidence presented for deliberative reflection. Nor do we mandate consensus among panelists by sequestering them, but rather permit a hung jury. However, the mathematical computation of an average or median standard to merge expectations among panelists can be as socially compelling in producing consensus as locking jurors in a hotel room.

Perhaps more troubling is the fact that standards panels seldom include a cross-section of socio-economic classes; the dynamic of stakeholder delegation processes regularly yields an educated, middle class, and predominately European-Canadian delegate — an inevitability the department actively counteracts by recruiting panelists from Aboriginal teacher training institutions. Yet are these not the interests who typically frame discourse around schooling in North America?

GENERALIZATION AND REPRESENTATION

*Trevor:* I do not believe that representativeness is the issue here. Clearly, a small jury cannot have perfect fidelity with the macroscopic community from which it is selected, because jury selection processes do not guarantee inclusion of all segments of society. What is sought, rather, is the involvement of informed individuals who come to the task with different perspectives and experiences, and without a strong political agenda, but driven by a firm mandate to be fair. This indeed is what transpired within my panel. No individual spoke directly to the position taken by her or his stakeholder organization. Since society as a whole is the beneficiary of the education of the next generation, then representation by defined groups within society will always be open to the question of why some groups are included and not others.

*Darryl:* Didn’t we have the same debate south of the border in the O.J. Simpson trial, where a jury of one composition found the accused not guilty, but another jury of a different composition in different circumstances with different purposes but identical evidence found the defendant liable? In fact, many segments of the American populace found the process highly political.

Those with grounding in statistics, with the aim of making the process objective and apolitical, believe that standard-setting panels should conform to the tenets of random probability sampling theory,
yielding a statistically reliable generalization of the total population’s attributes. Yet a statistically generalizable sample, for the audiences of a public accountability report in Saskatchewan, would have to number approximately one thousand. Pragmatic considerations overrule here; costs and logistics necessitate smaller samples. So we have to consider other alternatives in terms of the sampling strategy, so that it conforms to the purposes of the task at hand. Critics may assert that standards panels are formed by sampling delegates from the “politically important cases” (Patton, 1990), as a way of attracting attention to a study, as a means of promoting greater use of evaluation findings, or as subterfuge to avoid undesirable public scrutiny, by purposefully eliminating from the panel those cases which will be politically sensitive. In this optic, standard setting can become an exercise in garnering the votes of the powerful, articulating yet marginalizing the voices of the disadvantaged, and constructing a pseudo-consensus around those goals that authorities wish to remain front and centre.

I don’t subscribe to that sampling strategy. Ours might be better characterized as “critical case” sampling (Patton, 1990) — that is, choosing representatives from among the crucial organizational interests in the provincial educational setting. Here the purpose is not primarily political accommodation and reconciliation, but rather to permit logical extension of stakeholder values and to seek maximum application of information to other cases. In other words, standard-setting panelists are chosen not primarily because their judgments are statistically generalizable, nor because they represent an official organizational ideology, but because they reflect a value orientation within an organization.

Trevor: I am curious about your distinction between value orientation and ideology. Is that not just semantic splitting and dicing?

Darryl: I see values in philosophic terms as “concepts of the desirable with motivating force” (Hodgkinson, 1991, p. 89) or as considerations of what is important and not important. They are distinct from morals as conceptions of what is good or bad and from ethics which consider matters of right and wrong conduct. The hallmark of an ideology is its compelling moral vision that tends to bifurcate the world into endogenous believers and exogenous non-believers. An effective standard-setting panel will not easily accommodate those whose value positions have rigidified into ideological positions. When facilitating standard-setting exercises, on occasion, I have observed panelists lose their influence within group deliberations if
they attempt to turn the exercise into a soapbox for the “official line.” The effective panelist will function as a barometer of the organization’s values rather than solely as an amplifier of its ideology. As an evaluative exercise, standard setting thus serves as a vehicle for adducing, readjusting, and accommodating the different value systems among, and even within, key organizational interests.

The distinction between ideology and value position can be traced directly onto the sociograph of a stakeholder organization. I suggest drawing standard setters from outside the executive levels of organizations where values may have become fixed into a compelling moral or transrational vision through the dynamic of inter- and intra-organizational struggles for position. Standard-setting exercises may best be designed at a policy level through the interplay of executive ideologies, but best implemented with those who are influential in both shaping and refracting an organization’s values.

The recruitment of standard setters from those echelons of stakeholder organizations slightly below the executive level is what makes standard-setting exercises empowering in a sociological sense. Delegates will typically be “up-and-comers,” those who enjoy the confidence of the executive and who are upwardly mobile within that organization, but not yet in formal positions of executive authority. They have the ear of senior officials but are not yet exponents of an official ideology that is Manichean in its moral vision. Indeed, executive members are often pleased to devolve the task because they are too busy with other duties in their schedules. We must also remember that stakeholders generally prefer long-term influence over short-term advantage. As such, participation in a provincial standard-setting exercise often enhances the delegate’s stature within the sending stakeholder organization because she or he has functioned on behalf of the organization in an inter-organizational setting.

Trevor: There seem to be multiple ways of defining representativeness; the most traditional means extrapolative or “generalizable” in a statistical sense. That definition is one that Saskatchewan Education appears to have rejected in composing a standard-setting panel. Another, derived from a humanistic perspective, is “personificatory” in the sense that panelists give voice as “spokespersons” to a variety of positions within the educational system. It emphasizes collaborative approaches, consensus building, and the resolution of conflicting ideas about the purposes of schooling. That definition conforms to my approach. Yet another definition, and the one that I think underlies your motivation for involving stakeholders,
is more constructivist in orientation; a standard becomes “represen-
tational” because panelists stand in for roles and apply a mixture of
values from within and around the educational community to the
data. In effect, panelists serve as proxy readers of test information,
lending their personal, organizational, and hence social values to
the findings to create an interpretive framework for the findings.

CONCLUSION

Reflecting on our experience with educational standard setting in
Saskatchewan, we conclude that procedural fairness requires that
all participants have not only a voice but a vote. Divorcing a voice
from a vote raises strong emotion because the social dynamic of
empowering stakeholders cannot be separated from the power of
words. Within the framework of an inclusive and deliberative evalu-
ation process that serves democratic decision making (House &
Howe, 1999), authority derives not only from subject-matter expert-
tise but also from experience. In fact, the notion of power is more
complex in stakeholder approaches to evaluation than some writers
assert (Greene, 1988; Mark & Shotland, 1985). For example, char-
acterizing assessments as either “high-stakes” or “low-stakes” is mis-
leading, because it presumes an underlying, two-dimensional
continuum of “stakes.” In contrast, the notion of stakes is clearly
multivalent, differing not only according to the organization but ac-
cording to the individual. Stakeholder approaches to evaluation may
have an inherent tendency to oversimplify the variety of stakes at
play. This can lead to caricaturization of positions as postures, stere-
otyping of viewpoints, leading ultimately to dissension and perhaps
conflict.

In fact, there are both centrifugal and centripetal tensions within
stakeholder approaches to standard setting. On one hand, the
postmodern impulse to acknowledge diversity of perspective and to
respect a plurality of values and interests implies that consistency
need not be a prerequisite in collaborative, evaluative exercises. On
the other hand, the pursuit of a standard as a stable referent for
judging educational quality necessitates some convergence of val-
ues and a measure of consensual decision making. In Saskatchewan,
stakeholders generally concur that standards must be collaboratively
defined and are impelled to develop common indicators of educa-
tional quality, but are divided about the utility of standards. What
remains unexplored in standard setting is the degree of acceptable
divergence permissible within a stakeholder panel. A statistician
might ask, what is an acceptable standard deviation? For a policy
maker or educator, who must abide by the standard, the question
is: At what point does the standard disguise value positions so far
apart that the “collaborative” decision has little credibility among
the parties who made it?

Similarly vexatious issues arise when determining if a standard re-
fects idealistic expectations or estimated student abilities. In the
Saskatchewan exercise, participants were asked to consider and dis-
tinguish between “would” and “should” perspectives before making
their decisions. Yet, in practice, it was difficult to separate the two.
The moralist will point out a naturalistic fallacy in determining what
ought to be solely on the basis of what is. At the same time, we
detect a fallacy in assuming that teachers, chosen to bring a realistic
view of student abilities, will produce less idealistic standards
than non-educators. In fact, participants bring multiple perspectives
and values to an evaluative exercise, viewpoints that cannot easily
be ascribed to a particular organization. That is not to say that stan-
ard-setting exercises are as capricious as a bingo hall. Rather, it is
to say that an effective standard-setting exercise will reveal to par-
ticipants the complexity of educational roles and responsibilities and
demonstrate to those engaged the legitimacy of contrary positions,
with the ultimate goal that panelists will make reasonable adjust-
ments and balanced judgments.

Our notions of representativeness and generalizability are shaped
by what we see the standard as signifying or representing. In the
psychometric paradigm involving statistical generalization, the
standard signifies a mathematical relationship between samples and
larger populations. With an empowerment approach to evaluation,
the standard is more interpersonal or inter-organizational in mean-
ing; it reflects a power relationship. With a public administrative
decision-making rationale, the standard incarnates a values syn-
thesis. For those with subject-matter expertise, the standard often
represents conformity with academic norms, while for others reli-
ant on experience, it may signify aspirations for the future. Thus,
what is being generalized will vary from a projection of hopes, to an
extension of social and professional values, to an extrapolation of
scholarly conventions. We do know that standards do not drop in
tablet form from the heavens, but rather are created by humans
with their feet on the ground. Whether one engages stakeholders or
not, the question remains, “On whose ground are standards best
constructed?”
REFERENCES


Saskatchewan Education. (1999b). *Saskatchewan opportunities to learn related to reading and writing*. Regina: Author.


