Reforming Education: Is Inclusion in Standardization Possible?

Rosalyn Adamowycz, M.A.

University of Prince Edward Island

Abstract

Two reforms have evolved over the past fifteen years in the North American public education system, inclusion and large-scale assessment. The inclusion movement emerged from an educational reform to establish equal access to education, and the implementation of large-scale assessments stemmed from standards-led reform to encourage high standards for students. This article examines the implementation of these two complex educational movements; analyses how the large-scale assessment movement has incorporated inclusive practices; and presents existing examples that attempt to facilitate inclusive processes in large-scale assessment practices.

William Butler Yeats once stated, "Education is not the filling of a pail, but the lighting of a fire" (in Jones, 1999, p. 16). This quote is popularly used by educators and suggests that education ought to foster the love of learning in students rather than support the reiteration of rote memory. Development and planning in the public educational system in North America is riddled with diverging viewpoints. Rittel and Weber (in Friedmann, 1987) refer to the development of education as a "wicked" problem because no "black and white" solutions exist due to the multiple variables involved. Accordingly, a wicked problem is "'vicious' (like a circle) or 'tricky'" (Friedman, 1987, p. 166) meaning, problems and solutions are more obscure than transparent. Given this viewpoint the implementation of reforms in education is a many-sided venture.

This article examines the implementation of two complex educational movements, inclusion and large-scale standardized assessment, into the North American public education system. Separately, the implementation of the two educational movements has been laden with many reported advantages and disadvantages, which are summarized in this paper. In particular, inclusive practices require strong and extensive supports, and managing these supports within inclusive and large-scale assessments practices is a struggle. Subsequently, this paper presents an analysis on how the large-scale assessment movement has ineffectively incorporated inclusive practices, and proposes existing models that attempt to facilitate inclusive processes in large-scale assessment practices.

Background

Inclusion and large-scale assessment are two educational reforms evolved over the last fifteen years in the North American public education system. Reforms are often deemed as a movement of change away from traditional structures or processes. The inclusion movement emerged from an education reform to establish equity and equal access to education, and the implementation of large-scale assessments stemmed from a standards-led reform to encourage high standards for students. The inclusive model involves engaging and including students with varying abilities in the learning process; and the large-scale assessment model challenges students to improve academically. To refer back to Yeats's quote, it may be implied that the former model aims to light the fire

of learning, and the latter model aims to fill the pail of knowledge. Both models, however, are alike in that they aspire to enhance student learning and reform educational processes. This section describes the two models in terms of their purpose and practice. Further on, the two models are examined in terms of their co-existence with one another in the educational system and potential improvements.

Inclusion

As a reform, the inclusive model was influenced by the civil rights movement which addressed discrimination and segregation in society: "The civil rights movement had a major impact on society's attitudes towards segregation of minority groups. As a result of this socio-political perspective, students with disabilities began to receive increased attention" (Stewart, 2000, p. 11). The inclusive model can be referred to as a philosophical way of thinking, and in its purest sense, full inclusion implies equal access to education by learning the same curriculum in the same classroom for *all* students regardless of ability, learning style, behavioural issues, culture, or language. Schrag and Burnette (1994) characterize inclusive schools as a "philosophy of coordination that celebrates diversity and maintain a continuum of educational options to provide choice and meet the needs of individual children" (p. 2). The inclusive model aims to encompass a wide range of diversity; however, inclusion is most often practiced with students with disabilities. Although full inclusion means that "all children from all categories of disabilities and all degrees of severity [enter] into … regular education classes" (Lerner, 2000, 153), in some cases, this may mean taking choices away from students with severe special needs. Therefore, general educational practices do *not* follow a fully inclusive model as such. Providing an inclusive classroom setting for students with severe behavioural or developmental disabilities is challenging and may not be successful or, arguably, not appropriate in some cases.

The types of disability that inclusive classrooms address are those commonly determined by an Individual Education Plan (IEP). The IEP is produced by the teacher, parent, and other professionals to provide mandated supports. The IEP outlines curriculum accommodations, adaptations, or adaptive technology, and acknowledges special supports that may be needed. Some supports may be as simple as providing extra time for a student to process information and other supports may require visual

aids in the case of a student with visual impairment. Through the IEP, students with disabilities become included into regular classrooms. However, students with an IEP that *modifies* curriculum may not always be included in the regular classroom due to the severe nature of the disability (e.g. autism). A modified curriculum may veer from regular classroom outcomes extensively (Roeher Istitute, 2004).

Nevertheless, the inclusive model brought forth a break from the traditional, dualistic view of education where students were separated into regular and special education. Stewart (2000) describes the inclusive model as a departure from the isolation and discrimination against students with disabilities in the educational system. In addition, "inclusion demands that supports be brought to the classroom to the child, not that the child be removed to the supports" (Winzer, 1999, p. 44). This paper defines inclusion as equal access to education for students identified with disabilities, which involves bringing in supports into the classroom to address individual learning needs within the general education setting. It is acknowledged that some students with disabilities needing modified curriculum may require some supports outside the classroom. Regardless, the right to quality curriculum in teaching and learning remains as the goal in the inclusionary concept.

Clearly, the practice of inclusion in a classroom is complex; at one level all students are included, and on another level it is recognized that all students require different teaching strategies to learn. Winzer (1999) observes, "There is a general perception created by literature, and perhaps the media, that inclusion is a universally accepted movement in special education. This is not true. … [it] is both evolving and changing" (p. 40). Winzer further adds, "Most Canadian legislation speaks to the idea of inclusion but not to the practice" (p. 50). Educational ministries define and implement inclusion in a variety of ways across Canada (e.g. apply mainstreaming or integration as a variation of inclusion.) What the inclusive model strives for are classrooms where "… students work in flexible learning environments, with flexible curricula and instruction that are accessible to all. … what differs is the level at which these outcomes are achieved and the degree of emphasis placed on them" (Shrag & Burnette, 1994, p. 2). Correspondingly, to meet diverse learning needs, educational staff require diverse knowledge and specialized practitioners (Willrodt & Claybrook, 1995), as Schrag and

Burnette state "no one person can meet all student needs in a heterogeneous classroom" (p. 3). Parsons and Beauchamp (1995) suggest that a successful inclusive setting requires the following components: active parental involvement, technical support for faculty & family, individually designed integration plan, peer buddy program, ecological approach to curriculum, and creative management of behaviour problems.

An inclusive classroom is then based on a continuum of special education services for students with disabilities. Appropriately, inclusion requires the characteristics of diversity in all aspects of the educational process, including the staff and curriculum processes. An inclusive classroom requires an inclusive educational system and community.

Researchers and practitioners have observed various benefits and disadvantages in the inclusive model. According to Winzer's (1999) analysis, the advantages for students with disabilities in an inclusive classroom include: increased opportunity to interact with peers, meaningful curriculum content, positive role modelling, less stigma, and the fostering of positive attitudes. A positive learning environment through inclusion can foster greater confidence in students with disabilities and influence improved student achievement. Lerner (2000) states that the advantages in an inclusive classroom for students with learning disabilities include the "right to participate in environments as close to normal as possible and to benefit socially and academically …" (p. 148). This right includes having meaningful access to curriculum (adapted or modified) that strives for accountability and high expectations, with a focus on improved teaching and learning (IDEA '97, 1999).

The above advantages assume that inclusive classrooms are provided with resources for teachers and specialized staff to meet the individual needs of students with disabilities within the general classroom. Lerner (2000) admits that this model "… require[s] individual, intensive, explicit instruction that cannot easily be provided in the regular class" (p. 148). Without appropriate resources and planning Winzer (1999) observes that an inclusive classroom can have negative aspects including: detrimental effects on peer relations; untrained teachers who cannot manage students with varying disabilities, an atmosphere of ignoring differences, and the fact that one curriculum cannot teach all learning abilities. In fact, Winzer indicates "much research shows that

students with learning disabilities are poorly served in the general education classrooms" (p. 48).

Although educational policies are comparable across Canada, inclusionary policies and processes in education are established by each province and territory. The inclusive model does not have universal processes and variations exist across Canada. The implementation of supports is difficult to justify and access when "… there is not a consistent idea of what conditions should be considered when asking and answering questions about inclusion" (Parsons & Beauchamp, 1995, p. 13). In addition, some educational ministries may misconstrue the concept of inclusion to save education dollars, and thereby not provide supports needed for an inclusive classroom. For example, the American Federation of Teachers state "inclusion is being adopted by a large number of local school boards … as a means to save money be placing all students with disabilities in general education classrooms and curtailing special education supports and services" (in Verstegan & Martin, 1995, p. 8). In addition, Fuchs, Roberts, Fuchs, and Bowers (1996) examined the success of integrating students with learning disabilities in mathematics classes and commented on "… school administrators looking to full inclusion as a means of cutting costs at all costs (see Leo, 1994) – presents a probably more difficult scenario"(p. 31).

The philosophical premise behind inclusion is to accept the diverse ability of students into one classroom with appropriate supports to provide equal access to quality education. This reformed approach to education can encourage engagement, tolerance and confidence in all students, and subsequently positively influence learning and achievement for all students. The integration of this model in education is recognized by the fact that inclusive policies are implemented across all Canadian provinces and territories in varying degrees. Although the practices and implementation of the inclusive model are subject to refinement and evolution, the educational reform of integrating diversity is likely long lasting.

Large-scale Assessment

The large-scale assessment model stems from the American ideology referred to as standards-led reform in which it is assumed if educators "…set high standards for

student performance, develop assessments that measure student performance against the standards … then student achievement will rise" (Quenemoen, Lehr, Thurlow & Massanari, 2001, p. 1). As the researchers further explain, "As portrayed by the theory of action, the intended outcome of standards-based reform is increased levels of learning and achievement for all students in our nation's schools" (p. 2). Large-scale assessments are based on an action theory in which it is assumed that by setting high standards for students, student achievement increases as represented by the large-scale assessment scores. The large-scale assessment model is a process whereby all students are tested on prescribed academic topics at various grade levels.

Influenced by the American model, large-scale assessments in Canada arose from a belief "… that the educational system in Canada is no longer providing the level of education needed in today's society and that schools are failing" (Crudwell, 2005). Further, Crudwell observes "large scale assessments have become the vehicle of choice in the U.S. and Canada to address achievement and accountability". Assessments provide an uncomplicated definition of achievement and "suggest to the general public that the effectiveness of students, schools, and teachers can be moved into the world of statistics in which there is increased accuracy and objectivity in examining achievement" (ibid). In this model, the assessment standards are not compared to norms or a set of peers, but an outside standard set by educational ministries and statisticians. This is the essence of large-scale assessment practices across Canada. Linn and Herman (1997) explain, "…assessments compare student accomplishment to pre-established performance goals, rather than to the performance of other students. The standard is supposed to be absolute, independent of the proportion of students who meet it" (p. 6). This model assumes that all students should perform at par with an external established standard and that this standard reflects the attainment of student achievement. From this viewpoint, the "filling of the pail" is evidence of student achievement by way of quantitative scores.

Similar to the inclusive model, no universal or uniform procedures are in place and each province and territory implements its own form of large-scale assessment. For example, in Prince Edward Island (PEI), the last province to implement such an assessment, the province-wide test is termed a "Common Assessment" to signify its

connection to curriculum (PEI Department of Education, 2007a). Yet, PEI's "Common Assessment" uses outside standards to compare achievement, and therefore is essentially a large-scale assessment.

Although large-scale assessments are implemented across Canada, variations exist in terms of format, subject material, and process. Generally, Canadian jurisdictions design the large-scale assessments with questions based on the multiple choice format and essay or short-answer format. Other formats include "performance based" sections scored by selected provincial teachers, case studies, and observation. Across Canada, mathematics, reading and writing are assessed, however provinces and territories also include other subjects such as biology, chemistry, social sciences and history. In some provinces, a percentage of the assessment score is taken into account of the total grade in the school year. Overall, this model is widely utilized and as the Canadian Teachers Federation state, "Province-wide achievement testing of elementary and high school students is now the norm throughout Canada" (in Volante, 2004).

In practice, there are benefits to establishing a clear definition of student achievement and applicable measurements. The results of large-scale assessments can provide valuable information about curriculum and ways to improve student academic achievement. For example, Linn and Herman (1997) observe that assessments can communicate standards of achievement; provide achievement targets; and shape the performance of educators and students. Ensuring that students achieve a level of knowledge from the educational system to function in life is important to sustain a society - economically and socially. Certainly educational guidance and standards are useful otherwise students may become aimless in their goals. In this model, "…there is a focus on student achievement as the primary measure of school success… there is an emphasis on challenging academic standards that specify the knowledge and skills students should acquire … [and] there is a desire to extend the standards to all students,..." (Elliot, Braden & White, 2001, p. 8). As the researchers point out, "Meaningful information resulting from tests, however, can help teachers do their jobs better" (p. 7). For example, if student achievement is wanting in one academic area then curriculum can be examined and modified to improve learning outcomes.

Past research indicates various downfalls in the practice of large-scale assessments that appears to outweigh the benefits. In his analysis of the Ontario educational practices, Crudwell (2005) admits this model may show effectiveness and achievement in education but "as is often the case, statistics can be misused and misinterpreted". In his examination of validity and reliability information the author concluded, "… inherent weaknesses in current assessment practices reduce the ability to draw conclusions regarding accountability as expected by the general public". In an analysis of the design of large-scale assessments, the Canadian Association of Principals add "These invalid uses of large scale testing have been further exacerbated by the media, narrowly focused interest groups, or elected officials."

The assumption that quantifiable scores can demonstrate achievement may also be simplistic. For example, CBC news (2006) reported that the recent 2006 Ontario report on education indicated: "that despite the fact student test scores in Ontario have risen in comparison to others across Canada, the rate of students graduating actually fell to 71 per cent in 2004 from 78 per cent in 1995. Only Alberta's graduation rate is worse." One might assume that increased academic achievement scores would indicate the intelligence to remain in school. The article continues, "Another top concern is the number of students reporting they don't enjoy reading". In the CBC report the educational advocate and spokeswoman for People for Education who commissioned the report, Annie Kidder, stated "I'm worried that in our drive to set targets for test scores and focus on the mechanics of reading that we've lost the true meaning of literacy…". Is this an example of an educational system filling the pail, but not lighting the fire? Perhaps Yeats would say so. As Elliot et al. (2001) remind: "Tests do not create better students. Rather, good teachers and good schools do" (p. 7).

Rittel and Webber (in Friedmann, 1987) note that when the American system had the political dilemma to improve "the low average performance of high school students on standard achievement tests" (p. 167) the problem was deemed as "low scores" rather than considering other variables such as, lack of resources, poverty, broken homes, cultural adjustment and so forth. Therefore, the push to improve scores can be disguised as the "real" problem because this may be more easily implemented and

measured. Again, pointing to the notion that improving education or increasing student achievement is a layered or "wicked" problem with no simple solution.

Another effect of large-scale assessments is that it influences the curriculum and behaviours of both teachers and students to align with the testing instrument. This could be a positive outcome from one perspective. However, Cheng & Curtis (2003) refers to this effect of large-scale assessments as washback "… the extent to which the test influences language teachers and learners to do things 'they would not necessarily *otherwise* do because of the test'" (p. 6). Volante (2004) explains that in the American system where assessment results are tied to teacher salaries, institutional resources and status, educators have reverted to "teaching to the test" which leads to a "dumbing effect" by focusing on drilling practices, a narrow curriculum, and promoting memorizing rather than learning. When stakes are tied into the results of assessments, Cheng and Curtis (2003) remark on a number of negative aspects of washback including: a distorted curriculum whereby subjects not assessed are ignored; reduced instruction time for a wider curriculum; and a memorization approach to learning.

Although many criticisms are attached to the large-scale assessment model, this model shows itself as a popular initiative to improve student achievement results. If designed and implemented accurately, assessments may provide insight on how students are achieving and what areas require improvement. Assessment results may offer suggestions on what measures ought to be put in place to enhance student achievement from an academic viewpoint. While the belief in setting high standards for students to achieve is noble, it is crucial that the standards and the design of assessment instruments are a collective and transparent process that involves a diverse set of players (parents, students, educators, and professionals).

The two models – inclusion and large-scale assessment – represent educational reforms that have been incorporated in the Canadian educational systems for over a decade. Both models offer perspectives in enhancing student achievement. The inclusive model suggests that supporting diverse learning abilities in the regular classroom environment would foster engagement, confidence, and motivation, and consequently, enhanced student achievement. The large-scale assessment model suggests setting certain standards to motivate students to achieve well academically.

Both models vary in their implementation practices across Canada, and therefore are subject to evolve and develop. In addition, the models hold unique benefits and disadvantages as observed by researchers and educators. In practice, both models are currently implemented, which raises the question how do two models with diverging viewpoints in student achievement co-exist with one another? The following section explores this question based on the practises that have occurred, and offers potential avenues for further reform.

Analysis

Co-existence of Two Educational Reforms

As described, inclusion and large-scale assessment have been incorporated in the North American educational system to varying degrees for over a decade. Inclusion offers an avenue for all students to learn in a way that fosters engagement, equality and student achievement. The large-scale assessment model, as it pertains to provinces and territories in Canada, suggests setting certain standards to propel students to achieve well academically. Since large-scale assessment is a dominant focus and practice in the educational system, an examination is needed to determine the extent that large-scale assessment serves students with disabilities and the inclusion model. As Vinson (1998) inquires, "Does curriculum standardization imply anything in terms of cultural standardization? …the question is whether or not diversity can be standardized" (p. 30). Overall, the history of including students with disabilities into large-scale assessments has been dismal. Research on the co-existence of the two models raises doubts as to whether it is possible or practical to standardize diverse learning in terms of its design and implementation.

At its conception, the design of large-scale assessments did not account for students with disabilities. Before 1997, students with disabilities in America were purposefully excluded from large-scale assessments due to "…concerns about the test scores of students with disabilities lowering the overall scores of a school and concerns about the effect of assessments on the self-esteem or emotional health …"(Lerner, 2000, p. 100-101). Elliot et al. (2001) add that students with disabilities were excluded from assessments due to erroneous beliefs such as: protecting students from frustration or inequitable failure, deeming the assessment not relevant, and assuming

accommodations would jeopardize the validity to the test score. Although large-scale assessments are promoted to enhance teaching and learning, students with disabilities have often been excluded from being engaged in these gains. Regardless of the reasoning, Elliot et al. (2001) agree that exclusion led to unrepresentative scores, an incomplete picture of school achievement, beliefs that students with disabilities cannot accomplish work and an undermining of inclusion efforts. The large-scale assessment model has yet to provide a design that offers inclusion and flexibility for all students.

In 1997, legislation came in place through amendments to the Individuals with Disabilities Education Act (IDEA) to abide by inclusive practices and require students with learning disabilities to participate in assessments according to requirements and associated accommodations (Lerner, 2000). Therefore, large-scale assessments imply that "for all students, we must keep the standards high and do whatever it takes to help students be successful. There can be changes in the curriculum, the structure, the time it takes to learn, the way we assess, but there cannot be lowering of the standards" (Quenemoen et al. 2001). On one hand, it is theoretically suggested that large-scale assessments can by implemented for curriculum that is changed (adapted or modified). On the other hand, in practice, Almond, Quenemoen, Olsen, and Thurlow (2000) reveal a

> … disharmony between high standards and all students. All students are expected to reach high standards and the accountability system is being used to identify areas of curriculum and instruction that the schools must improve. However … large scale assessment systems seem lacking in their ability to assess and report what some students know and can do. These students include special populations such as students with disabilities, English language learners, and disadvantaged students."

One standard cannot assess curriculum that is adapted or occurs outside the regular classroom. Consequently, students with disabilities can potentially be excluded from large-scale assessment. Some Canadian provinces (e.g. PEI) exclude some students with disabilities, at the discretion of the school principal (PEI Dept. of Education, 2007b). The standards and the large-scale assessments themselves need to be flexible to integrate the inclusive model. This issue is complicated further in that the

inclusive model is not implemented consistently across Canada. How does a one-curriculum assessment instrument with arbitrary standards measure varied curriculum according to students' learning ability? A disconnect exists between reaching standards in an inclusive way.

Further, as noted previously, training for teachers on inclusive practices is often inadequate. Braden, Schroeder and Buckley (2001) argue, "…that if students with disabilities must be included in assessments to be counted in educational reforms, educators must have a conceptual framework for promoting inclusion and authenticity" (p. 1). Begging the question, are designers and administrators of large-scale assessments well informed in inclusion? In their 2003 analysis the Canadian Association of Principals (CAP) reveal, "The participation of disadvantaged groups (cultural differences, disability, language, access to community and family resources etc) is not sufficiently taken in to account in the preparation, delivery and interpretation of the test." A lack of knowledge about inclusion is inherent in the designs of large-scale assessments. Further, CAP state "Tests are often based on curricula whose learning outcomes have been only recently or poorly developed and communicated." The above statement presumes inadequate curriculum for 'regular' students and therefore one could only imagine the inadequacy of curriculum that addresses students with different learning styles and needs. As CAP observe, "Many large-scale tests currently in use in Canada do not adequately reflect or report on the variety of forms of intelligence and learning styles."

Another way inclusion can be tainted in the design of large-scale assessments is through the tendency to assign "stakes" onto achievement scores. For example, in the United States the stakes for assessment scores influence the amount of funding and resources a school may attain. This practice does not occur in Canada; however, large-scale assessment results are public and published. This type of public reporting often leads to certain positive or negative assumptions about certain schools and can effect an institution's reputation. Assigning stakes to assessment scores and public reporting can lead to school personnel to manipulate the exclusion option to increase overall scores and attain positive outcomes for the school (e.g. resources, status). In a study conducted by Figlio and Getzler (in Francis, 2003) the researchers discovered that

before the 1997 IDEA legislation "the schools did 'game the system' by reshaping the test pool. Schools reclassified students as disabled, … and therefore ineligible to contribute to the school's aggregate test scores. … schools reclassified low income and previously low performing students as disabled at significantly higher rates." The study noted an "increase in the likelihood that a student will be classified as disabled by 5.6 percentage points". Lehr and Thurlow (2003) agree that "… unintended consequences such as increased referrals to special education, low expectations for students with disabilities, and programmatic decisions based on incomplete or inaccurate information…" occurred out of large-scale assessment practices. The manipulation continued even after legislation was in place to include students with disabilities in the assessment process. The design of large-scale assessments provides "an incentive to place relatively high-achieving students, say those with mild dyslexia, into the disability category to improve the probability of attaining adequate yearly progress" (Francis, 2003). The pressure for schools to achieve a certain standard and reputation can lead to manipulation in the assessment process, usually at cost to the student, and with the result of inaccurate achievement scores and an ineffective policy.

The large-scale assessment process has not had a simple or consistent delivery system for students with disabilities. As previously stated, some Canadian jurisdictions exclude students with IEPs from the assessment process altogether. In the United States (where more publication about large-scale assessment and inclusion exists) it has been observed that students with disabilities may currently participate in the assessment process with supports in place, but how their scores are incorporated in the school calculations and statistics are often unclear (Bolt, Brentz, & Thurlow, 2001). In addition, other unintended consequences can occur via the IEP process with large-scale assessments. Quenemoen, et al. (2001) note that a student with an IEP can face lowered expectations by educators and thus receive lower outcome objectives on their plan to ensure that s/he masters the assessment. Further, these objectives may become misinterpreted achievement results.

Essentially, large-scale assessments have not been designed for all students, but focus on a restricted standard. Large-scale assessments have been designed for students who learn by status quo methods and have not exhibited the flexibility that the

inclusive model commands. Potential benefits exist for students with disabilities to participate in well-designed large-scale assessment, such as, guaranteed access to the general education curriculum, opportunity to learn and achieve at/near grade level, meaningful graduation steps, and accountability for all students (Quenemoen, et al. 2001). For all its attempts, the design and implementation of large-scale assessment has not effectively factored in the notions of inclusion or diversity in the learning process.

Facilitating Co-existence

Given that both inclusion and large-scale assessment are important initiatives in education, the current reforms call for ways to co-exist. Educational systems across Canada would benefit from responsibly designed assessments that account for the inclusionary model. A responsible assessment process would include a student's ability, socio-economic status, culture, language and equal curriculum access in learning. Ideally, the definition of student achievement would also be inclusive and not only include rote memory assessments (e.g. multiple choice format), but other forms of knowledge such as creativity, critical thinking or social intelligence. As Jones (1999) concludes in his analysis of the history of large-scale assessments "It is shortsighted to accept [academic] test scores as the ultimate criterion of the benefits of education, and more appropriate criteria must be developed." Education is a multi-leveled process and includes social, cognitive, emotional, cultural and developmental elements. In an analysis of how to improve student learning, Willms (2004) argues for schools that are learner friendly which includes building: a positive school atmosphere, good teacher relations, strong disciplinary climate, academic achievement, heterogeneous grouping, team teaching, and a high level of parent involvement. Therefore, emphasizing academics and providing an inclusive setting are part of a larger equation to improve learning. However, one missing element in the current large-scale assessment design is effectively accounting for inclusion. The following section reviews ways in which large-scale assessment might become inclusive, from re-designing current assessments to examining different assessment models.

To begin with, assessments must be clear as to what variables are being measured and with that ensure the assessment instrument is reliable. As Crundwell

(2005) notes "The current form [of large-scale assessment] in which most provinces release the results continue to lead to misinterpretation by many individuals and organizations, and they are often used to compare and rank schools even though they were not designed for these purposes". Often large-scale assessments purport to assess student achievement, system accountability, and shape teaching and curriculum. As concisely stated by Linn and Herman (1997) "An assessment that attempts to perform too many functions - … will inevitably do nothing well" (p. 23).

In the Policy Brief from the Atlantic Centre for Policy Research, Willms (1998) outlines ways in which large-scale assessment can be used to "constructively inform school practice and enhance student learning". Among the recommendations, Willms stresses the need to examine "non-cognitive" outcomes such as teacher-student relations, the learning and disciplinary atmosphere, and parental involvement. The design of the instrument must not only be comprehensive but also avoid negative side effects, such as, taking away teaching time in the classroom (i.e. teaching to the test) and publicly comparing schools (instead have an internal standard). Willms explains that an internal standard allows the school to examine their own process, while an external standard invites variables such as poverty and lack of resources to skew results. The assessment must also be flexible to account for different learning abilities and change accordingly. Essentially the "Assessment should stimulate critical inquiry and discussion about school renewal …." Willms explains this type of discussion includes input from the students and parents in the assessment process and results. Overall, a democratic and comprehensive assessment process is less likely to discriminate against students with disabilities, while at the same time informing ways that school processes could improve learning.

Additionally, CAP offers ten principles and eleven recommendations to re-design current assessment practices to minimize detrimental effects on student learning. As this is an exhaustive list, a few considerations are mentioned here. One recommendation suggests differentiating tests that are used for system accountability from student improvement. An instrument to access system accountability outcomes can be less intrusive, such as, performing a random sampling of student performances. Further, it is suggested to include variables such as socio-economic status and

community resources in the achievement formula and take the time to pilot and validate tests that are appropriate to diverse learning styles and abilities.

In the book, *Assessing One and All: Educational Accountability for Students with Disabilities* by Elliot et al. (2001) inclusive strategies for assessment are described. Inclusive options range from providing accommodations that do not alter the target skills tested, to providing alternative assessments for students requiring modified curriculum. The accommodations or modifications begin at the IEP process and are produced by a team "based on the individual needs of a student, not on the student's disability category" (p. 111). The authors suggest processes to ensure equal access to curriculum by matching the accommodations developed from the IEP to the content that will be taught and assessed. By establishing a system of assessment accommodations for students with disabilities it can be assured that all students are included in the same learning process.

Besides enhancing the current large-scale assessment design, other researchers suggest alternative assessment processes altogether. Grubb and Courtney (1996) argue that for younger students "assessments should be tailored to a specific purpose, but not to sort, sift, or label children" (p. 4). Because a young student is developing their cognitive processes and it is at this time when learning disabilities can be uncovered, the standard assessment process is artificial and "Provides little diagnostic information. Instead, it provides ranking information…" (p. 9). The researchers suggest the use of portfolio assessments whereby the work of the young student is gathered and assessed in terms of growth and learning. The portfolio model:

> demonstrate[s] growth over time, provide a means for student self-evaluation, help students determine and set individual goals, and provide real-life learning opportunities. Portfolios allow teachers to observe language development across ages and cultures, to evaluate and develop the curriculum, to determine efficacy of teaching practices, and to facilitate faculty discussion about goals and means" (6).

Crundwell (2005) also suggests an alternative assessment process called value-added assessment in which longitudinal assessment is administered to measure the individual student over time and accurately determine the development of achievement.

As Crundwell describes, this model originated from Dr. William L. Sanders who believed that learning was an individual process and not all students achieved simultaneously. The value-added model of assessing achievement and accountability is noted to resolve many of the current design flaws such as: relying on a single test to rank schools, not accounting for non-instructional variables, not accounting for different learning ability, and comparing to an arbitrary standard from a different population. Attempts to include value-added elements have been made in the past, but require commitment to a process that yields prolonged achievement results.

The above examples offer ways in which inclusion can be considered into the design and process of large-scale assessment. It is evident that the standards-led reform is not a perfect science and, in light of the drive to provide inclusion, large-scale assessment can be further developed and enhanced.

Conclusion

"Reforms in education are continuous and reflect a society's views of what is important at a given time" (Winzer, 1999, p. 33). The purpose of this article was to examine two models that grew out of the North American public education system. Inclusion and large-scale assessment are important in education although their co-existence has not been successful. Large-scale assessment has led to some detrimental effects for students with disabilities. The design of this assessment model tends to confuse outcomes: "When the system is held accountable for students' performance, there typically is a push to not include students with disabilities and other students considered to be low performing" (Thurlow, 2000). Potentially, large-scale assessment may become inclusive when the time is taken to create a comprehensive and responsible instrument. Various researchers (Crundwell; Elliot et al.; Grubb & Courtney; Willms) offer possibilities for large-scale assessments to become inclusive. Perhaps one main struggle of the inclusive model is that it requires an inclusive educational system and society. As Winzer (1999) observes, "maintaining such needed support is an enormous task that requires high degrees of commitment, communication, cooperation, collaboration, and funding" (p. 44).

The goal to enhance student learning for all students and ensure that schools are providing adequate services is a society's responsibility: the policy makers, educators,

administrators, parents and students. Goals such as inclusion and large-scale assessment can facilitate the process of enhancing student achievement, but must be implemented comprehensively. Inclusion without support and resources does not succeed. Large-scale assessment without transparency, equality, validity and reliability does not succeed. However, reforms that integrate both equality and a comprehensive assessment of knowledge may contribute to a meaningful education for the future.

References

Almond, P., Quenemoen, R., Olsen, K. & Thurlow, M. (2000). *Gray areas of assessment*

      systems. (Synthesis Report No. 32). University of Minnesota: National Center on

      Educational Outcomes. Retrieved 11 April 2007 from

      http://education.umn.edu/NCEO/OnlinePubs/Synthesis32.html.

Bolt, S., Brentz, J.& Thurlow, M. (2001). *Are we there yet? Accountability for the performance*

      of students with disabilities. Technical Report. University of Minnesota: National

      Center on Educational Outcomes. Retrieved 30 Nov 2007 from http://www.

      Education.umn.edu/NCEO.

Braden, J.P., Schroeder, J.L.& Buckley, J.A. (2001). *Secondary school reform, inclusion, and*

      authentic assessment. Brief: Research Institute on Secondary Education Reform

      (RISER) for Youth with Disabilities. (ERICDigests No. ED467477).

Canadian Association of Principals (CAP). (Feb. 2003). *Valid uses of student testing as part of*

      authentic comprehensive student assessment. CAP Position Statement: CAP.

      Retrieved 10 May 2007 from

      www.cdnprincipals.org/CAPStatementonTestingarticle.pdf.

CBC News, Toronto. (May 10, 2006). *High school graduation rate in Ontario below 1995*

      levels: Report. Toronto: CBC News Web site. Retrieved 10 May 2006 from

      http://www.cbc.ca/toronto/story/tor-education060510.html.

Cheng, L., & Curtis, A. (2003). Washback or backwash: A review of the impact of testing

      on teaching and learning. In L. Cheng & J. Watanabe with Curtis, A. (Eds.)

      Washback in language testing: Research contexts and methods. New York:

      Lawrence Erlbaum Associates, Inc.

Crundwell, M.R. (2005). Alternative strategies for large scale student assessment in Canada: Is

value-added assessment one possible answer. *Canadian Journal of Educational Administration and Policy, 41.* Retrieved 11 April 2006 from www.umanitoba.ca/publications/ cjeap/pdf%20files/crundwell.pdf.

Elliot, S.N., Braden, J.P, & White, J.L. (2001). *Assessing one and all: Educational accountability for students with disabilities.* Arlington, VA: Council for Exceptional Children.(ERICDigests No. ED458746).

Francis, D.R. (2003). *How school administrators cheat the accountability rules.* Report: National

Bureau of Economic Research. (NBER Working Paper No. 9307) Retrieved 16 April 2007 from http://www.nber.org/digest/jun03/w9307.html.

Friedmann, J. (1987). *Planning in the public domain: From knowledge to action.* Princeton, N.J.:

Princeton University Press.

Fuchs, D., Roberts, H., Fuchs, L.S., & Bowers, J. (1996). *Reintegrating students with learning disabilities into the mainstream: A two-year study.* Conference Paper: Annual Convention of the Council for Exceptional Children. (ERICDigests No. ED385052).

Grub, D. & Courtney, A. (1996). *Developmentally appropriate assessment of young children:*

The role of portfolio assessments. Conference Paper: Southern Early Childhood Association. (ERICDigest No.ED400114).

IDEA '97. (1999). The role of general educators: Linking to the general education curriculum

and participation in state- and district-wide assessment. *Issues in Education. Technical Assistance Bulletin.* (ERIC Digests No. ED436897).

Jones, L.V. (1999). *The assessment of student achievement: The hundred years war.* Conference Paper: Annual Meeting of the American Educational Research Association. (ERICDigests No. ED430009).

Lehr, C. & Thurlow, M. (2003). *Putting it all together: Including students with disabilities in*

assessment and accountability systems. University of Minnesota: National

Center on Educational Outcomes. (Policy Directions No. 16). Retrieved 15 April

2007 from http://education.umn.edu/NCEO/OnlinePubs /Policy16.htm.

Lerner, J.W. (2000). *Learning Disabilities: Theories, diagnosis, and teaching strategies* (8th

ed.). New York: Houghton Mifflin Company.

Linn, R.L. & Herman, J.L. (1997). *Standards-led assessment: Technical and policy issues in*

measuring school and student progress. University of California: CRESST. (CSE

Technical Report 426). Retrieved 2 Mar 07 from

cse.ucla.edu/Reports/TECH426.PDF.

Ministry of Education, Ontario Government. (2006). *Struggling students to get new help:*

intensive small-group tutoring and literacy camps. Ontario Government Web site.

Retrieved 11 April 2007 from http://ogov.newswire.ca/ontario

/GPOE/2004/04/08/c9723.html?lmatch=&lang=_e.html.

Parsons, J. & Beauchamp, L. (1995). *Teaching in an inclusive classroom: An essay to young*

teachers. Report: University of Alberta. (ERICDigests No. ED391794).

PEI Department of Education. (2007a). "Assessment". Presentation by C. Wood, Manager,

Assessment and Evaluation Unit. University of Prince Edward Island. Oct 31, 2007.

PEI Department of Education, C. Wood, Manager, Assessment and Evaluation Unit. (business

email communication Dec 2007b). Providing information on the Common

Assessment exclusionary regulations for PEI students with disabilities.

Roeher Institute, The. (April 2004). Inclusive Policy and Practice in Education: Best Practices

for Students with Disabilities. Report. Retrieved 30 Nov 2007 from

http://inclusiveeducation.ca/documents/BEST-PRACTICES.pdf.

Quenemoen, R. F., Lehr, C.A., Thurlow, M.L., & Massanari, C.B. (2001). *Students with*

disabilities in standards-based assessment and accountability systems: Emerging issues, strategies, and recommendations. University of Minnesota: National Center on Educational Outcomes. (Synthesis Report 37). (ERICDigests No. ED452654).

Schrag, J. & Burnette, J. (1994). Inclusive schools. *Research Roundup,10(2), 1-4.*

Stewart, L.M. (2000). *Connecting regular and special education pathways at the high school level:. Case studies on inclusive education.* Unpublished doctoral dissertation, University of Alberta, Edmonton.

Thurlow, M.L. (2000). *Including special needs students in standards-based Assessments: Special needs students in standards-based reform.* Report: McREL's Diversity Roundtable. Retrieved 12 May 2007 from www.mcrel.org/PDF/Diversity/5007IR_DiversityRT3.pdf.

Vinson, K.D. (1998). *National curriculum standards and social studies education: Dewey, Freire, Foucault, and the construction of a radical critique*. Paper Presented: Annual Meeting of the National Council for the Social Studies. (ERICDigests No. ED438194).

Volante, L. (2004). Teaching to the Test: What Every Educator and Policy-maker Should Know. *Canadian Journal of Educational Administration and Policy, 35.*

Willrodt K & Claybrook S. (1995). *Effects of inclusion on academic outcomes*. Unpublished research report. Texas, USA: Sam Houston State University. (ERICDigests No. ED389102).

Willms, J.D. (2004). *Raising and leveling the bar*. Policy Brief: Atlantic Centre for Policy Research, UNB. Retrieved 7 March 2006 from www.unb.ca/crisp/pbrief.html.

Willms, J.D. (1998). *District and Provincial Assessment: Can it be Useful to Principal And Teachers?* Policy Brief: Atlantic Centre for Policy Research, UNB Retrieved 7 March 2006 from www.unb.ca/crisp/pbrief.html.

Winzer, M. (1999). Issues and trends in Canadian special education. In *Children with Exceptionalities in Canadian Classrooms. (*5th ed.) Scarborough: Prentice Hall Allyn and Bacon Canada, 31-65.