CHAPTER NINE

ASSESSMENT^{15}

In Chapter Eight, we saw how the trend in western countries was for SWSEN to participate and progress in the general curriculum, albeit with appropriate modifications and adaptations. In this chapter, parallel issues will be explored with respect to assessment, namely the extent to which SWSEN are expected to participate in a country’s national or state assessment regimes and what, if any, alternate assessment procedures are permitted. Both trends are part of the wider concern for standards-based reform in education that is dominating much of the educational and political discourse around the world^{16}. The vast bulk of literature on modified and alternate assessment has emanated from the US and this section of the review reflects that.

9.1 Policies Requiring Access to General Education Accountability Systems

*United States.* Until recently, in the US, accountability in special education was defined in terms of progress in meeting IEP goals. That all changed in IDEA 97, which required all students, including those with disabilities, to participate in their states’ accountability systems. This was followed by a policy memorandum from the U.S. Department of Education (2000), to the effect that an exemption from a state’s assessment programmes was no longer an option for students with disabilities. Both IDEA 97 and the No Child Left Behind Act (NCLBA) of 2002 required the provision of alternate assessment for students who could not participate in state or district assessments with or without accommodations. Districts are permitted to measure up to 3% of their students using alternate assessments (1% against alternate achievement standards and 2% against modified standards – a distinction that will be described in more detail below). The use of alternate assessment is a decision to be made by a student’s IEP team. To quote IDEIA, IEPs must include ‘a statement of any appropriate accommodations that are necessary to measure the academic achievement and functional performance of the child on state- and district-wide assessments’ (IDEIA, 2004, p.118). As well, the NCLBA stipulated that student performance be disaggregated by special education status, among others, and, to avoid sanctions, by 2013/2014 schools must show that students in

^{15} This chapter is mainly drawn from Mitchell et al. (2010) and Mitchell (2008).
^{16} See Chapter Six, section 6.5.
various subgroups are making adequate yearly progress toward mastering content standards.

At this juncture, it is worth quoting at length a personal communication from David Egnor, Assistant Division Director, National Initiatives, Research to Practice Division, Office of Special Education Programs, US Department of Education:

… one of the main pushes in the U.S. particularly among special education administrators, but also teachers, is to develop standards-based IEPs. I believe that standards-based IEPs are becoming much more attractive from an administrative point of view as a direct result of our country's increasing focus on standards-based educational reform … and which will ratchet up even further under the Obama administration. That is, requiring standards-based IEPs for every student with a disability (not currently required for all students with disabilities, although things are moving that way) provides a way, from an administrative perspective, to more efficiently administer and monitor special education service delivery and to do so within a standards-based accountability environment, where, in the past, special education practice historically focused more on individualized services and outcomes for students with disabilities. My view is that the growth of standards-based IEPs in the U.S. is a clear sign that special education practice is undergoing fairly significant changes that are directly tied to standards-based reform under the ESEA/NCLB and the next iteration of our main federal education law currently under consideration in the US Congress. I think that what we are seeing with regard to standards-based IEPs is an outgrowth of the special education inclusion movement, where as a field special education attempts to make the general education environment more accessible to students with disabilities. Given the focus on standards-based educational reform, it is not surprising that special education administrators, in particular, seek a way to join with the standards-based movement through the IEP development process and, as a result, students' IEPs are emphasizing general education standards more and more. Although a standards-based IEP should not limit the services a student receives (just standardize, to some extent, the educational outcomes we expect), I think that this movement may be unintentionally limiting services for some students with disabilities. I also think that more work needs to be done to explicate how individualization (equity) for students with disabilities can co-exist within the growing context of standards-based reform (excellence).

According to Defur (2002), the thinking behind the earlier requirements was two-fold. Firstly, it was assumed that higher expectations would lead to higher achievement for students with disabilities. Previously, the educational progress of such students had been limited by low expectations, which in turn narrowed their access to the general curriculum and to higher achievement. The second assumption was that assessment information on students with disabilities would lead to improved instructional programmes, which in turn would lead to improved student outcomes. It would seem that this rationale still applies.
England and Wales. In England, tasks and tests set for assessment at the end of Key Stages 2 and 3 (for students aged 11 and 14, respectively) are designed to monitor attainment targets for each of the National Curriculum subjects, and are expected to be accessible to the vast majority of students, including those with special educational needs. However, those children in Key Stage 2 working at level 1 or below of the National Curriculum eight-level scale are assessed by teacher assessment alone. Similarly, at Key Stage 3, students working at or below level 2 of the National Curriculum scale are assessed by teacher assessment and not by statutory national testing. If a student's statement of special educational needs modifies the statutory assessment arrangements, the provisions within the statement should be followed in respect of the statutory tests and tasks. With regard to the GCSEs and GCE A levels, although the same examinations are available for SWSEN as for other students, special arrangements in examinations may be made for some of them. The nature of these arrangements is determined according to the assessment needs of the individual student, but must not give him or her an unfair advantage over other students. Some may be awarded extra time to complete the assessment task, or may be permitted to take supervised breaks or rest periods during the examination. For visually impaired students, the visual presentation of the papers may be changed by, for example, the use of large print or simplified layout of the examination paper, or by the use of braille versions of the papers. Other candidates may have questions read to them; flashcards may be used to assist hearing-impaired candidates in mental arithmetic tests; or typewritten, word processed or transcribed responses may be accepted from students who are unable to write. Some candidates may also be allowed to take their examinations at a venue other than the examination centre, for example, at home or in hospital (see http://www.inca.org.uk/wales-sources-special.html#31).

In England, too, the ‘P Scales’, referred to in Chapter Eight, can also be employed to provide a means of assessing students with special educational needs for accountability and school improvement purposes, prior to them becoming eligible for assessment on national instruments. These P Scales have eight levels against which students’ progress can be mapped. However, Riddell et al. (2006) while recognising that P Scales are helpful for curriculum planning, noted that ‘whether they will be useful in terms of tracking and comparing the progress of pupils with special educational needs has yet to be fully assessed’ (p.5).
Scotland. According to Riddell et al. (2006), in Scotland there are ‘ongoing difficulties in devising a national system of assessment which is able to recognise the progress of all pupils’ (p.5). The Standard Grade system, they pointed out, is regarded as too difficult for some students with special educational needs, particularly those with significant difficulties in numeracy and literacy.

9.2 Adaptations, Modifications and Alternate Assessment

Geenen & Ysseldyke (1997) identified six types of accountability systems relating to the extent to which students with disabilities are included in assessment regimes:

*Total inclusion.* This type establishes a single set of standards, with one assessment programme for all students, including those with disabilities. At the time of writing [1997], two US states had developed portfolio-assessment programmes that covered all students.

*Partial inclusion.* Here there is one set of standards for all students, with alternate or modified standards for students with disabilities. Many states were adopting this arrangement.

*Dual systems.* This type involves two sets of standards: one for students without disabilities and another one for students with disabilities, the latter usually focussed on ‘functional’ objectives.

*Multiple systems.* Here there is one set of standards for students without disabilities and multiple sets of standards for those with disabilities, usually based on their disability category.

*Total exclusion.* In this type, students with disabilities are excluded from standard-setting efforts, state-wide assessments, and data-based reporting procedures. Usually, the IEP is seen as sufficient for accountability purposes, despite the difficulty in aggregating their outcomes.

*System-based.* This sets standards on a system rather than an individual basis. Here, students with disabilities ‘count’ in the overall statistics.

Research relating to one or more of the models as outlined by Geenen & Ysseldyke (1997) has been reported in the literature.

For example, in a paper by Defur (2002), the Virginia state assessment programme was outlined. This state employed the *total inclusion model,* albeit with accommodations/modifications/exemptions in parts of the tests for students with disabilities (the author pointed out that after her study, Virginia eliminated the use of total exemptions). It is interesting to note that 98 special education administrators in the state identified some intended and unintended consequences of this assessment policy. Among the intended consequences were (a) ‘some degree of benefit for students with disabilities’ - reported by 83% of the respondents, (b) ‘access to the general curriculum’
(73%), and (c) ‘improved daily performance by students with disabilities’ (but only 21% noted this) (p.206). There were also unintended, negative consequences of the policy. These included (a) higher rates of academic failure (reported by 51% of the administrators), (b) lower self-esteem among students with disabilities (50%), and (c) concerns that these students would experience higher drop-out rates (44%). As well, some were of the opinion that standards should be lowered (33%) and that accommodation options should be increased (37%). And, finally, 55% of the respondents expressed the belief that special education teachers were not adequately trained to assist students with disabilities to meet Virginia’s assessment standards.

In full inclusion assessment models, with no exemptions or accommodations permitted, there is a risk that ‘the accountability procedures may have the incidental effect of discouraging schools from taking on children who are likely to perform poorly in examinations, of encouraging schools to expel children whom they find difficult to teach, or of tempting schools to omit children with learning difficulties from testing programmes’ (OECD, 1999). As proof of this danger, OECD cited a study by Thurlow in 1997 in which it was found that two-thirds of students with disabilities in US schools had been excluded from a National Assessment of Educational Progress. Thus, ‘high stakes’ assessments, and associated ‘league tables’ can have the effects of jeopardising inclusive education (Dyson, 2005; Sle, 2005; McLaughlin & Jordan, 2005). As Watkins & D’Alessio (2009) pointed out, this risk can be exacerbated by the effects of international comparative studies of educational standards – most notably OECD’s PISA studies.

A second study, involving the partial inclusion model, was reported by Browder et al. (2004). Subject specialists and experts in severe disabilities from 31 US states were surveyed and interviewed regarding their views on the extent to which alternate assessment content was aligned with academic and functional curricula in maths and the language arts. The findings were quite mixed, with some states rated as having a high degree of alignment and some having missed the mark. The authors also noted that their results suggested that the alternate assessments included in their study had a strong focus on academic skills, but also reflected an approach that linked academic and functional skills, one which they referred to as ‘a blended curriculum approach’ (p.221). Browder et al. concluded with the recommendation that states should include both content area specialists and experts in severe disabilities in validating performance
indicators used in alternate assessment. In another paper by the same authors (Browder et al., 2003), some lessons to be drawn from their research are outlined. These included the need to develop research into (a) ways of teaching students with severe disabilities the more advanced academic skills that were being expected under the US legislation, (b) the impact of alternate assessment in general, and (c) the optimal way of blending functional and academic curricular priorities, and hence assessment approaches. And, finally, they argued that ‘We also need to avoid a transformative approach in which academics become the replacement curriculum’ (p.179).

In a similar vein, Ford et al. (2001) posed some pertinent, albeit rhetorical, questions. Firstly, when a state develops separate standards for students with disabilities, is it suggesting there is no overlap between the 98% of the students included in the regular assessment and the 2% who are not? Secondly, when states elect to use identical standards for those participating in alternate assessment, ‘does this mean that all students should be held to the same set of standards – and that these are the only valued areas of learning?’ (p.215).

In another US study involving Geenen & Ysseldyke’s (1997) partial inclusion model, Ketterlin-Geller et al. (2007) investigated the consistency of test accommodations across 38 3rd grade students’ IEPs, teachers’ recommendations, and students’ performance data. They defined accommodations as representing ‘changes in the medium through which information is presented, the response formats, the external environment, or the timing of the testing situation that are designed to mediate the effects of a student’s disability that inhibit understanding or expression of domain-specific knowledge’ (p.194). They found significant differences among all three of the comparisons, i.e., students’ IEPs, teachers’ recommendations, and students’ performance data. For example, individual teachers often made accommodation decisions without support from the IEP team and there was little correspondence between the accommodations listed on IEPs and teacher recommendations. As Ketterlin-Geller observed, ‘IEPs were more likely to make errors of omission, whereas teachers were more apt to make errors of commission in recommending accommodations’ (p.203). With respect to the latter errors, the researchers commented that by making decisions without recognition of the IEP, teachers may be subverting the legal requirements and that this may significantly affect student success by withholding accommodations or by providing unnecessary accommodations. This, they concluded,
compromises both students’ needs and the accountability systems set up to ensure that their needs are being met. ‘The current system’, they stated, ‘needs improvement’ (p.205).

In yet another US study, Karnoven & Huynh (2007) investigated the relationship between IEP characteristics and test scores on an alternate assessment instrument for students with significant cognitive disabilities. They found that whereas the curriculum emphasised in IEPs and alternate assessments were aligned for some students, for others they were not. They concluded that teachers of such students, who may have operated outside the general education curriculum for many years, ‘need professional development on state academic standards, alternate achievement standards, and curriculum design that goes beyond functional domains’ (p.291). As well, they argued that there is a need to create standards-based IEPs and that test developers must contribute to improving the curriculum-assessment link.

For other studies of alternate assessments and some attendant concerns, see papers by Browder et al. (2003); Crawford & Tindall (2006), Kohl et al. (2006), NAREM Associates, in cooperation with OECD (2005), Rabinowitz et al. (2008), Salend (2008), Thompson & Thurlow (2000), Turner et al. (2000), and Zatta & Pullin (2004).

In the US, the National Center on Educational Outcomes has published extensively on alternate assessment for students with significant cognitive disabilities (see Lazarus et al., 2010a and 2010b; Olson, et al., 2002; and Quenemoen et al., 2003). These documents are too lengthy to summarise here, but suffice to say they provide information on States’ accommodation policies on alternate assessments and guidelines for such assessments. Other useful guides to alternate assessment are to be found in the recently published book by Bolt & Roach (2009) and in publications from the US Department of Education, particularly those relating to its policy for including students with disabilities in standards-based assessment used in determining ‘adequate yearly progress’ (Technical Work Group on Including Students with Disabilities in Large Scale Assessments, 2006).
9.3 Some Definitions of Assessment Accommodations and Alternate Assessments

Basically, there are two types of adjustments to nation- or state-wide assessments.

Assessments with accommodations. This involves making changes to the assessment process, but not the essential content. Braden et al. (2001) described accommodations as alterations to the setting, timing, administration and types of responses in assessments. Here, assessors need to distinguish between accommodations necessary for students to access or express the intended learning content and the content itself.

Alternate assessments. As defined by the US Department of Education (2003), alternate assessments are defined as assessments ‘designed for the small number of students with disabilities who are unable to participate in the regular State assessment, even with appropriate accommodations’ (p.68699). They refer to materials collected under several circumstances, including: teacher observations, samples of students’ work produced during regular classroom instruction, and standardised performance tasks. Further, alternate assessments should have:

- a clearly defined structure,
- guidelines for which students may participate,
- clearly defined scoring criteria and procedures,
- a report format that clearly communicates student performance in terms of the academic achievement standards defined by the State, and
- high technical quality, including validity, reliability, accessibility, objectivity, which apply, as well, to regular State assessments.

Quenemoen et al. (2003) provided more detailed definitions and examples of the following alternate assessment approaches:

*Portfolio*: a collection of student work gathered to demonstrate student performance on specific skills and knowledge, generally linked to state content standards. Portfolio contents are individualized and may include wide ranging samples of student learning, including but not limited to actual student work, observations recorded by multiple persons on multiple occasions, test results, record reviews, or even video or audio records of student performance…

*IEP-Linked Body of Evidence*: Similar to a portfolio approach, this is a collection of student work demonstrating student achievement on standards-based IEP goals and objectives measured against predetermined scoring criteria…This evidence may meet dual purposes of documentation of IEP progress and the purpose of assessment.
**Performance Assessment:** Direct measures of student skills or knowledge, usually in a one-on-one assessment. These can be highly structured, requiring a teacher or test administrator to give students specific items or tasks similar to pencil/paper traditional tests, or it can be a more flexible item or task that can be adjusted based on student needs. For example, the teacher and the student may work through an assessment that uses manipulatives and the teacher observes whether the student is able to perform the assigned tasks….

**Checklist:** Lists of skills, reviewed by persons familiar with a student who observe or recall whether students are able to perform the skills and to what level. Scores reported are usually the number of skills that the student is able to successfully perform, and the settings and purposes where the skill was performed.

**Traditional (pencil/paper or computer) test:** Traditionally constructed items requiring student responses, typically with a correct and incorrect forced-choice answer format. These can be completed independently by groups of students with teacher supervision, or they can be administered in one-on-one assessment with teacher recording of answers.

For useful descriptions of alternate assessments for students with significant cognitive disabilities, see Perner (2007), who gave examples of various States’ methods, such as portfolio and performance-based assessments referred to above.

### 9.4 Formative Assessment

As might have become apparent in the foregoing, there is a tension between the need for schools to ascertain students’ level of achievement for accountability purposes and the need to take account of what is best educationally for SWSEN (Bauer, 2003). This distinction is sometimes referred to ‘assessment of learning’ (or summative assessment), compared with ‘assessment for learning’ (or formative assessment) (Harlen, 2007; Watkins & D’Alessio, 2009). If the purpose is to compare students against predetermined standards, then the former is best suited; if the purpose is to improve learning, the latter should be used.

Mitchell (2008) has summarised the distinction between summative and formative assessment. Briefly, *summative* assessment is concerned with evaluating learners’ performances at the end of a module or a course. The results count towards making a final judgement on what the learners have achieved. *Formative assessment* evaluates students’ progress during a course or module so that they have opportunities to improve, and teachers to ‘fine tune’ their teaching. In its pure form, formative assessment does not contribute to the overall grade. However, sometimes assessment serves both summative and formative purposes. How one classifies the two types depends on the extent to which
assessment leads to feedback that enables learners to improve their performances. The more it does this, the more justified is its classification as formative assessment.

There is evidence to suggest that formative assessment has a positive effect on learning outcomes for SWSEN. Three US studies will serve as examples of such research. Firstly, in an early meta-analysis of 21 studies of the effects of formative evaluation, an effect size of 0.70 was obtained. However, when formative evaluation was combined with positive reinforcement for improvement (i.e., feedback), the effect size was even higher at 1.12 (Fuchs & Fuchs, 1986). Secondly, a study using formative evaluation system with low-achieving students in a large urban school system resulted in significant gains in math achievement (Ysseldyke, 2001). Thirdly, there is evidence to show that teachers trained in formative assessment are more open to changing their instructional strategies to promote learners’ mastery of material (Bloom et al., 1992). Furthermore, it has been shown that without formative assessment, teachers’ perceptions of learners’ performances are often erroneous (Fuchs et al., 1984).

Finally, in a related vein, in recent years, the European Agency for Development in Special Needs Education has argued that assessment processes can either contribute to or hinder the process of inclusion (see various documents on the Agency’s website: www.european-agency.org). Thus, it has focused on what it refers to as ‘inclusive assessment’, which it defines as:

an approach to assessment in mainstream settings where policy and practice are designed to promote the learning of all pupils as far as possible. The overall goal of inclusive assessment is that all assessment policies and procedures should support and enhance the successful inclusion and participation of all pupils vulnerable to exclusion, including those with SEN (Watkins, 2007, p.47).

Educational policy-makers, then, should optimise both the needs of the system and those of its students in determining assessment policies.

9.5 Functional Behavioural Assessment

In the US, a major variant of the IEP is the ‘Behavior Intervention Plan ‘(BIP), with its reliance on ‘Functional Behavior Assessment’ (FBA). BIPs came into force in the US with the 1997 reauthorisation of IDEA, and were reiterated in the 2004 IDEIA. As described by Killu (2008) and Etscheidt (2006), BIPs consider the relationship between student learning and any behaviour problems they manifest that may impede their classroom performance or that of other students. A point of distinction between IEPs
and BIPs is that the latter must not only focus on individuals, but must also address school-wide issues that serve as contextual factors that may contribute to the behavioural problems (Killu, 2008).

In a review of FBA, 22 studies focused on learners with or at risk for emotional and behavioural disorders were reported. These studies comprised a mix of antecedent-based interventions, consequence-based procedures and a combination of the two interventions. Regardless of the type of intervention, 18 of the 22 studies showed positive results, with clear reductions of problem behaviours and/or increases of appropriate behaviours (Heckaman et al., 2000).

The principles of FBA are not limited to behaviour, but in recent years have been extended to learning difficulties as well (Daly & Martens, 1997; Jones & Wickstrom, 2002; Duhon et al., 2004).

9.6 Summary

1. Increasingly, SWSEN, including those with significant cognitive disabilities, are being expected to participate in their countries’ national or state assessment regimes.
2. High stakes' assessments can have the effects of jeopardising inclusive education, a risk that can be exacerbated by the effects of international comparative studies of educational standards.
3. In the US, legislation since IDEA 1997 does not allow SWSEN to be exempted from their states’ assessment programmes. Instead, educational authorities are required to provide alternate assessment for students who cannot participate in state or district assessments with or without accommodations. IEPs now must include a statement of any accommodations that are necessary to measure the academic achievement and functional performance of such students on state- and district-wide assessments.
4. The main types of alternate assessments comprise portfolios, IEP-linked bodies of evidence, performance assessments, checklists and traditional paper and pencil tests.
5. The assumptions underlying these provisions are twofold: (a) that higher expectations will lead to improved instructional programmes and (b) that these will lead in turn to higher student achievement.
6. The requirements for all students to participate in state- and district-wide assessments have been shown in some research to have had unintended negative consequences for students with disabilities, including higher rates of academic failure, lower self-esteem, and concerns that they would experience higher drop-out rates.

7. Countries or states should include both content area specialists and experts in severe disabilities in validating performance indicators used in alternate assessment.

8. With the shift to all students being required to participate in their countries’ national or state assessment regimes, teachers of SWSEN will need professional development on their country’s or state’s academic standards, alternate achievement standards, and curriculum design that goes beyond functional domains.

9. Formative assessment has been associated with positive outcomes for SWSEN and with improvements in teachers’ perceptions of students’ performances.

10. Functional assessment is increasingly being applied, not only to behaviour, but also to learning in general.

11. In determining assessment policies, it is important to recognise and resolve as far as possible the tensions between measuring the health of the education system and protecting the interests of students with special educational needs. In other words, educational policy-makers should optimise both the needs of the system and those of its students in determining assessment policies.
CHAPTER TEN

EVIDENCE-BASED PEDAGOGY

Educators are increasingly expected to be responsible not only for helping students to achieve the best possible outcomes, but also for using the most scientifically valid methods to achieve them. Indeed, in the United States, the No Child Left Behind (NCLB) law requires teachers to use ‘scientific, research-based programs’, defined as: ‘(1) grounded in theory; (2) evaluated by third parties; (3) published in peer-reviewed journals; (4) sustainable; (5) replicable in schools with diverse settings; and (6) able to demonstrate evidence of effectiveness.’ As well, NCLB requires each state to ensure that all learners (including those with disabilities) make ‘adequate yearly progress’, i.e., ‘continuous and substantial improvement’.

In their recent review of special education in the ACT, Shaddock et al. (2009) proposed ‘increased accountability for the learning outcomes of students with a disability and the adoption of evidence-based policy to inform service development’, arguing that ‘data and evidence, not conviction and ideology, are the key considerations’ (p.16). In a similar vein, the President’s Commission (2002) in the US recommended the establishment of ‘long-term programs of research that support evidence-based practices’ (p.61).

Briefly, evidence-based teaching strategies may be defined as ‘clearly specified teaching strategies that have been shown in controlled research to be effective in bringing about desired outcomes in a delineated population of learners’ (Mitchell, 2008, p.1).

10.1 Do SWSEN Require Distinctive Teaching Strategies?
The answer to this question is both ‘Yes’ and a qualified ‘No.’ Firstly, yes: some students – especially those with high or very high needs – do require some significantly different teaching strategies to those that educators in regular classes might usually employ. For example, some students with visual impairments are reliant on their tactile and auditory senses for learning and will require specialised techniques such as Braille and orientation and mobility training. Secondly, no: for the most part, SWSEN simply require good teaching. As some writers argue, there is little evidence to support the

17 This chapter is based on Mitchell (2008 and 2009).
notion of disability-specific teaching strategies, but rather that all learners benefit from a common set of strategies, even if they have to be adapted to take account of varying cognitive, emotional and social capabilities (Kavale, 2007). What is required is the systematic, explicit and intensive application of a wide range of effective teaching strategies (Lewis & Norwich, 2005).

Although they all have a substantial evidence base for SWSEN, almost all them have general applicability.

10.2 Criteria for What Constitutes Evidence

Ideally, evidence that a particular strategy works should be based on carefully designed research studies that meet criteria such as the following:

*Treatment fidelity*. The teaching strategy is fully described and there is evidence that it has been carefully implemented.

*Behavioural outcomes*. The study should include reliable and valid measures of the behavioural outcomes. When he selected the teaching strategies described in his recent book, Mitchell (2008) relied heavily on various meta-analyses that have been reported in educational literature. Briefly, a meta-analysis synthesises the results from a range of similar research studies to determine the average effect of a particular intervention. Meta-analyses usually produce a numerical indicator, known as *effect size*. The larger the effect size, the greater is the impact of the intervention. An effect size of 1.0 indicates that learners receiving the intervention would achieve better than 84% of those who did not receive it; an effect size of 0.7 means that those receiving the intervention would do better than 76% of those who did not; an effect size of 0.3 means scores better than 62%, and so on. Most of the strategies selected by Mitchell had effect sizes between 0.3 and 0.7, with some over 1.0.

*Learner characteristics*. Studies should include clear descriptions of the learners’ ages, developmental levels, and the nature and degree of any disabilities they may have. Ideally, research studies should focus on learners who are as homogeneous as possible. The more heterogeneous the sample studies, the more difficult it is for educators to decide which learners would benefit from the strategy.

*Control of variables*. The research should be designed to ensure that the outcomes are due to the intervention and not to any confounding variables such as the simple passage of time or a placebo effect. One would also want to be confident that the
outcomes are not due to the effects of additional attention to the learners in the study or to the effects of repeated testing.

*Freedom from contamination.* There should be no, or minimal, ‘contamination’ which might affect the results of the study. In other words, it is important that nothing happens (outside of the intervention) that could affect the outcomes for either the experimental group or the control group. Of course, if events occur that affect *both* the experimental and the control groups, that is acceptable.

*Acceptable side effects.* Possible side effects should be assessed and should be positive, or at least not negative. For example, coercive means might be used to control certain learner behaviours, but they may cause heightened anxiety or even fear.

*Theory-based.* The psychological mechanisms or learning processes underlying the strategy should be clearly explained, thus enabling one to generalise it to other situations.

*Follow-up.* There should be been adequate follow-up after, say, six months, but preferably longer, to ascertain if the behavioural gains are maintained over time.

*Research versus natural conditions.* Ideally, the research should be carried out in everyday teaching environments, not just in research conditions. This is because it could well be that the research conditions are dramatically different from the actual conditions educators work in.

*Peer review.* The research should have been published in reputable journals after rigorous peer review.

*Replication.* The research should contain at least two studies (more for single-case studies) that have shown positive effects for the strategy; i.e., the research has been replicated, preferably by independent researchers.

*Cost effectiveness.* Clearly, for an intervention to be adopted it must not be excessively expensive. For example, the more the intervention depends on one-to-one treatment over a prolonged period, the less likely it is considered to be cost effective.

### 10.3 Evidence-based Teaching Strategies

By applying as many as possible of the above criteria, Mitchell (2008) arrived at a total of 24 strategies, some of which included several sub-strategies. Although they are
illustrated with reference to learners with special educational needs, almost all the strategies have general applicability.

Mitchell emphasised that he was not arguing for a single strategy or blueprint that all teachers should use. Rather, he felt that the most effective programmes are those that incorporate a variety of best practices. His strong advice was that educators should develop a repertoire of such strategies, nested within their own philosophy, personality, craft knowledge, professional wisdom, and, above all, their knowledge of the characteristics and needs of their students and their knowledge of local circumstances.

In this chapter, 18 of the strategies are presented (Mitchell, 2009). They are as follows:

1. cooperative group teaching
2. peer tutoring
3. review and practice
4. formative assessment
5. feedback
6. cognitive strategy instruction
7. self-regulated learning
8. memory strategies
9. reciprocal teaching
10. behavioural approaches
11. social skills instruction
12. positive, motivating classroom environment
13. adequate active learning time
14. information and communications technology
15. parent involvement
16. phonological processing
17. optimal physical environment
18. combined strategies

In the following selection only a single representative study will be cited for each strategy. For a full review of the evidence, see Mitchell (2008).

*Co-operative group teaching.* A comprehensive study researched the effects of co-operative learning on the reading achievement of elementary students with learning disabilities. A total of 22 classes with 450 3rd and 4th grade learners, including those with learning disabilities, were involved in the study. Teachers in nine of the classes used an approach called Co-operative Reading and Composition (CIRC) to foster comprehension and metacognitive strategies. The other 13 classes formed the controls. In the CIRC classes learners worked in heterogeneous groups on activities including partner reading, examining story structures, learning new vocabulary, and re-telling
stories. Significant results were reported in favour of those in CIRC classes on standardised reading and writing tests (Stevens et al., 1987).

**Peer tutoring.** In a study of the effects of peer-assisted learning strategies on students’ reading achievement in 22 U.S. elementary and middle schools, 20 teachers implemented the programme for 15 weeks and 20 control teachers did not. It was found that all three groups of learners (low achievers with and without disabilities and average achievers) demonstrated greater reading progress (Fuchs et al., 2002).

**Review and practice.** In a comprehensive meta-analysis of 93 intervention studies targeting adolescents with learning disabilities, the single most important strategy was found to be explicit practice, defined as ‘treatment activities related to distributed review and practice, repeated practice, sequenced reviews, daily feedback, and/or weekly reviews’ (Swanson & Hoskyn, 2001).

**Formative assessment.** In an early meta-analysis of 21 studies of the effects of formative evaluation, an effect size of 0.70 was obtained (Fuchs & Fuchs, 1986). (See also Chapter Nine of the present review).

**Feedback.** After synthesising a large number of studies on the effects of a wide range of influences on learner achievement, a Hattie (2003) found 139 that focused on feedback. With an effect size of 1.13, this was the most powerful of all the influences on achievement. He concluded that ‘The simplest prescription for improving education must be ‘dollops of feedback’ – providing information how and why the child understands and misunderstands, and what directions the student must take to improve’. Although Hattie’s meta-analysis was not confined to SWSEN, it is highly likely to apply to such learners.

**Cognitive strategy instruction.** In a Canadian study of 166 learners, aged seven to 13 years, with developmental reading disabilities, three groups were identified: (a) those with deficits in phonological awareness, (b) those with deficits in visual naming speed (i.e., word recognition speed), and (c) those with both deficits. A metacognitive phonics programme resulted in improvements, especially for learners with only phonological deficits. This programme instructed the learners in the acquisition, use, and monitoring of four word identification strategies. These included, for example, a ‘compare/contrast’ strategy in which the learners were taught to compare an unfamiliar word with a word they already knew (Lovett, et al., 2000).
Self-regulated learning. In a meta-analysis of 99 studies that used interventions to decrease disruptive classroom behaviour, self-management strategies yielded an effect size of 1.00. In other words, there was a reduction of disruptive behaviour for about 85 per cent of the students treated by this method (Stage & Quiroz, 1997).

Memory strategies. Several research studies have shown that students (including those with a range of disabilities) can be trained to use memory strategies independently across a range of different content areas. For example, in 19 meta-analyses of various interventions, mnemonic training, with an effect size of 1.62, was rated the highest. This effect size can be translated to mean that the average student receiving mnemonic instruction was better off than 95 per cent of the students not receiving such instruction (Mastropieri & Scruggs, 1989).

Reciprocal teaching. A New Zealand study investigated the efficacy of a tape-assisted reciprocal teaching programme, referred to as ‘cognitive bootstrapping’ (LeFevre, Moore & Wilkinson, 2003). The study’s subjects were learners aged from eight to 10 years with poor comprehension skills, half of whom also had poor decoding skills. The results showed that the poor decoders improved their use of cognitive strategies and their comprehension. These results were maintained after 10 weeks and transferred to other material.

Behavioural approaches. In a comprehensive review of meta-analyses involving 20 different intervention strategies, behaviour modification came out with the third highest effect size (after mnemonic strategies, reading comprehension and just ahead of direct instruction). The effect size of 0.93 for behaviour modification represented the average of effect sizes for social outcomes (0.69) and academic outcomes (1.57) (Forness, 2001).

Social skills instruction. A recent UK study found that two social skills training interventions directed at primary school learners at-risk for social exclusion had positive effects on their social skills and social inclusion (Denham et al., 2006).

Positive, motivating classroom environment. A Dutch study found that educators who were perceived to be understanding, helpful and friendly and show leadership without being too strict, enhanced learners’ achievement and affective outcomes. Those who were seen as being uncertain, dissatisfied with their students and admonishing were associated with lower cognitive and affective outcomes (Wubbels et al., 1991).
Adequate active learning time. An Australian investigation found that when ‘wait-time’ was greater than 3 seconds, changes in educator and learner discourse took place and higher cognitive level achievement was obtained in elementary, middle and high school science classes. This finding was attributed to both educators and learners having additional time to think (Tobin, 1987).

Information and communications technology. A recent review of the literature on the use of computer assisted instruction (CAI) with learners with mild and moderate disabilities found that, although mixed, research supported the potential for CAI to raise academic achievement, particularly when it is used as a tool for extended practice of previously learned concepts (Fitzgerald & Koury, 1996).

Parent involvement. A 1996 meta-analysis of the effects of behavioural parent training on anti-social behaviours of children yielded a significant effect size of 0.86 for behaviours in the home. There was also evidence that the effects generalised to classroom behaviour and to parents’ personal adjustment (Serketich & Dumas, 1996).

Phonological processing. An Australian study evaluated the effects of phonological processing skills training for learners aged nine-14 years with persistent reading difficulties. The results showed that improvement in the learners’ phonological processing skills led to considerable improvement in their reading accuracy and reading comprehension. Extending the length of the training time significantly improved the transfer of skills to the reading process, especially for those with severe phonological processing skill difficulties (Gillon & Dodd, 1997).

Optimal physical environment. A New Zealand study examined the effects of sound-field amplification for four learners with Down syndrome aged six to seven years. The results showed that the learners perceived significantly more speech when the system amplified the investigator’s voice by 10 dB (Bennetts & Flynn, 2002).

Combined teaching strategies. A few studies have investigated the impact of two or more teaching strategies on learners’ academic achievement and social behaviours. Many of them have combined cognitive strategy instruction with another type of intervention, including direct instruction (Swanson, 2000), information and communications technology (Woodward & Rieth, 1997), phonological training (Lovett et al., 2000), and co-operative group teaching (Swanson, 2000). One Canadian study

---

18 See also Chapter Fourteen of the present review.
looked at the combination of three strategies: co-operative group teaching, teacher collaboration and parent involvement (Saint-Laurent et al., 1998).

10.4 A Scale for Evaluating Teachers’ Use of Evidence-based Strategies

In his recent paper presented at a UNESCO conference, Mitchell (2009) outlined a scale for evaluating teachers’ use of the strategies outlined above. The scale is designed to be used in carrying out a needs analysis for teachers’ professional development. This could involve the following three steps:

*Step One.* Teachers are asked to complete a questionnaire, rating their use of the 22 key strategies. The questions are intended to provide a broad picture only and provide a basis for a more detailed analysis to be conducted in the next step.

*Step Two.* This step would normally involve an independent evaluator who would build on a teacher’s questionnaire responses and would use a combination of an in-depth interview, classroom observations and document inspection to evaluate the teacher’s use of the 22 strategies. Mitchell noted that it might be possible for some teachers to carry out a self-evaluation of their use of the strategies, thus obviating Step One.

*Step Three.* On the basis of information obtained in the previous two steps, a professional development programme is designed.
EXCERPT FROM A SCALE FOR EVALUATING STRATEGIES FOR ENHANCING LEARNING (DRAFT)
©David Mitchell, 2009

NB: This Scale has yet to be peer-reviewed and tested for reliability. It should not be used until these steps have been taken and a revised form provided. Readers of this draft are invited to provide comments.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Indicators</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Employs co-operative group teaching</strong></td>
<td>1. In most lessons the teacher uses co-operative group activities.</td>
<td>A. All the indicators are regularly met.</td>
</tr>
<tr>
<td>The teacher regularly uses co-operative group teaching in which all learners work together in small learning groups of 6 to 8, helping each other to carry out individual and group tasks. Groups are usually mixed ability, but are sometimes comprised of learners with similar ability. The teacher teaches group process skills and carefully supervises group interactions.</td>
<td>2. The teacher uses a combination of (a) mutual assistance groups in which learners are encouraged to help individuals to carry out tasks, and (b) ‘jig-saw’ type groups in which all learners contribute to a group task.</td>
<td>B. The teacher occasionally uses both forms of co-operative group activities with ability groups and mixed ability groups.</td>
</tr>
<tr>
<td><strong>2. Employs peer tutoring</strong></td>
<td>1. In most lessons the teacher uses peer tutoring.</td>
<td>C. The teacher occasionally uses mutual assistance groups.</td>
</tr>
<tr>
<td>The teacher regularly sets up peer tutoring in which one learner (a ‘tutor’) provides learning experiences for another learner (a ‘tutee’). Such tutoring is mainly used to promote fluency through practising or reviewing skills or knowledge. The tutors are taught to follow a structured lesson format. Each dyad works for no more than 10 minutes at a time for 8-10 sessions.</td>
<td>2. The peer tutoring is used for practice and review of previously taught material.</td>
<td>D. None of the indicators are met.</td>
</tr>
</tbody>
</table>

10.5 A Final Word

The overarching theme of this chapter is that teaching must become more based on empirical evidence of what has been proven to be effective strategies for improving students’ outcomes. A secondary theme is that, in order to bridge the research-practice gap, it is necessary that teacher education - both pre-service and in-service must be
upgraded to deliver programmes based on evidence (see also Chapter Thirteen). Only by doing this will teaching be able to lay claim to being a true profession.

**10.6 Summary**

1. *Educators are increasingly expected to be responsible not only for helping students to achieve the best possible outcomes, but also for using the most scientifically valid methods to achieve them.*

2. *Evidence-based teaching strategies may be defined as ‘clearly specified teaching strategies that have been shown in controlled research to be effective in bringing about desired outcomes in a delineated population of learners’.*

3. *All students, including SWSEN, benefit from a common set of strategies, even if they have to be adapted to take account of varying cognitive, emotional and social capabilities. What is required is the systematic, explicit and intensive application of a wide range of effective teaching strategies.*

4. *To constitute evidence, research studies should meet criteria such as the following: (a) treatment fidelity, (b) reliable and valid measurement of behavioural outcomes, (c) adequate control of variables, (d) freedom from contamination, (e) adequate follow-up, (f) replicated in more than a single study, and (g) cost effectiveness.*

5. *Strategies that have a strong evidential base for use with SWSEN (and other students) include (a) cooperative group teaching, (b) peer tutoring, (c) formative assessment, (d) feedback, (e) cognitive strategy instruction, and (f) instruction in memory strategies.*

6. *A scale for evaluating teachers’ use of evidence-based teaching strategies is described.*

7. *In order to bridge the research-practice gap, it is necessary that teacher education - both pre-service and in-service must be upgraded to deliver programmes based on evidence.*