

Why National Standards must allow for the Special Character of Schools

Sebastian Suggate

PhD Candidate, Dept of Psychology,
University of Otago, NZ

(Written in June 2009, when submissions were being gathered by the Government concerning the introduction of annual National reporting of standardised testing from the first year of school till aged 16 in New Zealand- with thanks to the author for his support of the Steiner School stance –ed).

The government of New Zealand has recently passed legislation requiring that schools regularly assess their students and compare the test results to national standards. The aim is to “lift achievement in literacy and numeracy (reading, writing, and mathematics) by being clear about what students should achieve and by when” [1]. The stated goal, therefore, is that comparisons to national standards would formatively facilitate children’s progress – however, as will be argued, national standards become invalid if schools have a special character.

Without doubt, test development and psychometric evaluation has been the focus of a great deal of psychological and educational research, resulting in some quite impressive achievements [e.g., 2]. However, there are also misconceptions about what test results mean, such that indiscriminate provision of test results may—particularly in the case of schools with a special character—cause confusion and alarm, thus hindering the role of the school.

A score on a standardised test simply means that on that day, at that particular moment, the score that a particular child received, places him or her at a certain rank on the test normative sample. During psychometric validation of tests, statistical relationships between groups of scores on the particular measure are compared with how those particular children scored on other measures. This provides evidence of the validity for tests (i.e., do they measure what they are supposed to measure) and reliability evidence is obtained from, for example, retesting the same children on an alternate-form of the same measure and seeing if their scores are consistent.

There are, however, two characteristics of standardised tests that should be mentioned. First, the psychometric evaluation depends largely on the performance of not

individual children, but groups of children. Second, the statistical relationships between scores on the particular measure and the comparison test-validation measure are never perfect (in fact a strong relationship would be considered if a correlation coefficient in the vicinity of $r = .60$ were obtained between two tests of the same skill. This would mean that only 36%—the square of $.60$ —of the variance in measure A is predicted by measure B; in short there will be a significant proportion of children who perform comparatively poorly on one test and comparatively better on another measuring a similar skill). Therefore, in fact, as the Ministry of Education caution, scores for individual students should not be over-interpreted – evidence about how children are achieving needs to be gathered from multiple sources.

The most important consideration in the use of standardised testing is that the children’s scores are compared to the normative sample. In the case of schools with a different curriculum, the psychometric *validity* and *utility* of such a comparison may evaporate entirely.

To take a specific example, children do not begin explicit learning of reading skills in Waldorf schools until they are older, around seven years. By contrast, many children on the day of their fifth birthday in New Zealand enter state schooling and begin learning to read. Most state schools do—and have done so for years—report quarterly progress on semi-standardised tests to the Board of Trustees (e.g., Running Records) and, at age six, most schools administer the six-year-net [3] to a significant proportion of their students. From our research at the University of Otago, where we have been (a) evaluating measures of early literacy development [4, 5], and (b) investigating the reading development of children in state and Waldorf schooling [8], let me present some data to shed some light on this.

We have a measure of nonsense word reading that can be given to children in the middle of their first year of schooling and we have found that this measure predicts later reading quite strongly both for state ($r = .65$) and Waldorf ($r = .72$) school children. However, if the scores on this measure are compared against an estimate of national

norms (i.e., the state school performance) it would be found that the Waldorf pupils were in the 7th percentile in the middle of Class I, compared to age-matched state school pupils at the end of year 2. Parents receiving *only* this piece of information meaning that, on average, their children are coming 93rd in a race of 100 people, might be quite alarmed. However, as evidence now converges to suggest—including data from New Zealand—their children will not be disadvantaged by the later start in reading, once they are a few years older [6-10]. Therefore, the comparison to non-Waldorf norms, provided little useful information about “what students should achieve and by when” [1] – because they go on to achieve well despite the seemingly poor initial score.

This brings me to my conclusion – there is nothing to fear about the results of standardised assessment, because these *may* provide useful information, of more concern is if the data are not interpreted correctly. If it were the case that state requirements were such that standardised testing were required, then Waldorf parents would need to be informed that the results for particular children at certain ages were invalid, because they applied to an irrelevant normative sample, based on a pedagogy with an entirely different philosophy.

Clearly, comparing children’s scores on some standardised measures to the national norms would not be psychometrically sound practice, and may even be considered unethical if it creates undue concern and confusion. Therefore, adoption of national standards must respect schools’ special character.

References

1. **Ministry of Education.** *National standard factsheets: Reporting to parents.* 2009 [cited 2009 30 June]; Available from: http://www.minedu.govt.nz/the_Ministry/Consultation/NationalStandards/Factsheets.aspx.
2. **Fuchs, L.S. and D. Fuchs,** *Monitoring student progress toward the development of reading competence: a review of three forms of classroom-based assessment.* *School Psychology Review*, 1999. **28**: p. 659-671.
3. **Clay, M.M.,** *An observation survey of early literacy achievement.* 1993, Auckland: Heinemann.
4. **Schaughency, E. and S. Suggate,** *Measuring basic early literacy skills amongst year 1 students in New Zealand.* *New Zealand Journal of Educational Studies*, 2008. **43**: p. 85-106.
5. **Suggate, S.P. and E.A. Schaughency,** *The development of reading in Year 2 and its relation to reading in Year 3.* *The Bulletin*, 2007. **109**: p. 40-42.

6. **Durkin, D.,** *A six year study of children who learned to read in school at the age of four.* *Reading Research Quarterly*, 1974-1975. **10**: p. 9-61.
7. **Schmerkotte, H.,** *Ergebnisse eines Vergleichs von Modellkindergärten und Vorklassen in Nordrhein-Westfalen [Results from a comparison of typical kindergartens and preschools in Northrhine-Westphalia].* *Bildung und Erziehung*, 1978. **31**: p. 401-411.
8. **Suggate, S., Schaughency, E., and Reese, E.,** *Age of beginning formal reading instruction and later literacy: evidence from longitudinal research.* in *International Congress of Psychology*. 2008. Berlin, Germany.
9. **Suggate, S.P.,** *School entry age and reading achievement in the 2006 Programme for International Student Assessment (PISA).* *International Journal of Educational Research*, 2009: p. doi:10.1016/j.ijer.2009.05.001.
10. **Turnbull, S.,** *The rate of learning to read in a Rudolf Steiner school.* *Journal for Waldorf/Rudolf Steiner teachers*, 2006. **8**(2): p. 11-16.

Further References

- Amrein, A.L. & Berliner, D.C.** (2002, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved [date] from <http://epaa.asu.edu/epaa/v10n18/>.
- Audrey L. Amrein, A. L., & Berliner, D. C.** (2002) An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing. Arizona State University The Great Lakes Center for Education Research & Practice PO Box 1263 East Lansing, MI 48826 Phone: (517) 203-2940 Email: greatlakescenter@greatlakescenter.org Web Site: <http://www.greatlakescenter.org> December 2002
- Barksdale-Ladd, M. A., & Thomas, K.** (2000). What’s at Stake in High-Stakes Testing. *Journal of Teacher Education*, 51(5), 384-397. DOI: 10.1177/0022487100051005006
- Paris, S. G. Lawton, T. A., Turner, J. C., & Roth, J. L.** (1991) A Developmental Perspective on Standardized Achievement Testing. *Educational Researcher*, 20(5), 12-20.
- Shepard, L. A.** (1992). Will National Tests Improve Student Learning? Report - abstract available through ERIC
- Sloane, F. C., Kelly, A. E.** (2003) Issues in High-Stakes Testing Programs *Theory into Practise*, 42(1).
- Smith, M. L.** (1991). Put to the Test: The Effects of External Testing on Teachers. *Educational Researcher*, 20(5), 8-11
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System*, 28(4), 499-509.