

# **An Advocacy for New Placement Methods**

by  
Steven M. Snyder

## **Placement, Aptitude and Diagnostic Testing**

This article will be concerned with placement testing in academic language learning environments and with how developments in other areas of testing can improve placement tests. Hughes (1989) defines placement tests simply as "...intended to provide information which will help to place students at the stage of the teaching programme most appropriate to their abilities." However, Hughes is careful to note that "ability" is meant in its most general sense, and not in the sense of "talent" or "aptitude." Bachman (1990) states that placement testing can involve a variety of tests for a variety of purposes and uses, including diagnostic testing,

Information from language tests can be used for diagnosing student areas of strength and weakness in order to determine appropriate types and levels of teaching and learning activities... A placement test can be regarded as a broad-based diagnostic test in that it distinguishes relatively weak from relatively strong students so they can provided learning activities at the appropriate level. (Bachman 1990, p. 60).

Is placement testing an activity that involves testing for language aptitude and for diagnosis? Or should placement testing be thought of as distinct from other testing activities? The giving of placement tests in foreign language education is often justified as an effort to separate those students who learn languages more easily from those who learn languages more slowly. Is that enough? What are the consequences of these divisions? Is the relative ease of learning foreign language a sound basis for grouping students?

I will argue in this article that simply dividing students into groups is not an effective way to approach placement testing and that there is inherent in such an approach a significant opportunity cost to language programs in terms of lost information that would be of

use in improving instruction and student growth. I will also be arguing for the use of aptitude testing measures for placement and for diagnostic reasons. Further, I will argue that placement tests, like aptitude tests, must accurately predict learner performance, not just evaluate prior learning in the second language.

Perhaps for many language educators who are faced with placing a large number of students in a program, the distinctions between whether a particular test is measuring a native ability or if it is measuring the current mastery of the language, or both, may be irrelevant: all that may be desired in these circumstances is a test which is easy to grade, easy to administer and separates incoming students into groups. Such a cavalier attitude is, in my opinion very dangerous. It is important that we have valid tests; that is, that our tests actually are measuring the knowledge or skills that we claim to be testing. Validity can also be thought of in terms of how adequate the assessment is and how appropriately the results are applied.

One working distinction between placement testing and aptitude testing, which is only indirectly stated in the literature, seems to be that an aptitude test taps some native ability for learning languages, whereas placement testing is thought to be concerned with placing students by their current abilities or a combination of these two factors. As a practical matter, we cannot ignore the level of current language skill that a student has in making placement decisions. As Bachman (1990) has pointed out, the content of a placement test should depend upon circumstances and curricular requirements. Additionally, previous learning behaviors, such as course grades, are often a fair predictor of future learning performance (Carroll, 1979). On the other hand, in academic language learning environments experience teaches us that ranking only by current skill level is a less than satisfying way to place students. Firstly, the majority of students entering an academic program tend toward a similar skill level, unless they are from very differing backgrounds. In programs which simply divide students into two skill groups, those scoring in the middle range will be grouped half with the upper-performing students on the test and half with the lower-performing students on the test. In programs which are able to group students into three groupings, students in the upper middle will be placed with the highest skilled students and the reverse with the lower middle, regardless of how different the skill levels are between those groupings. For the rare programs which can make more groupings, the problem may be less pronounced, but a focus on only current skill levels would ignore other potentially important factors affecting future performance in the program.

A counter-argument to the foregoing claims that greater precision for placement testing is unnecessary because students who fall near the cutoff points would benefit equally whether

they were placed in an upper group or a lower group -- only gross misplacement, which is unlikely, is of any concern. This counter-argument does have considerable merit, if we assume three things about the placement test: 1. that the test is accurately predicting later learning performance; 2. that the test given has not excluded important factors which would affect performance; 3. that the learning in both groupings would be equivalent and not adversely affected by the placement procedure itself. I do not believe that the majority of placement tests within college and university programs would satisfy these assumptions. In the development of this article I will attempt to make clear just why this is the case.

A second objection to the use of current knowledge of the L2 as a sole criteria for placement is that it may not be a good predictor of student learning performance in the language program. It is not uncommon to encounter students who appear to know more of the L2 than their peers, but do not improve much during their studies; conversely, there are those students who did not have the benefit of much exposure or instruction in the L2 prior to entering the program whose learning performance within the program may be very good. An obvious explanation for this variance is that there are other factors which affect learning performance than just previous knowledge of the L2.

### **What is foreign language aptitude?**

There is little agreement on the precise definition of aptitudes, but for foreign language aptitude the discussion has centered on the relative ease with which a person learns a foreign language. It is a commonly observed phenomenon that some students do indeed learn more easily than others (Ellis, 1994; Wesche 1979; Carroll, 1979). Presumably, there are varying aptitudes for learning languages that account for the differences in how easily students learn languages. J.B. Carroll noted that in his work in language aptitude research he had,

...been impressed with the observation that variations in success manifested themselves chiefly in rate of learning; it thus seemed a natural conclusion to draw that aptitude should be defined in terms of prediction of rate of learning. (Carroll 1979, p.91)

Such aptitudes are often thought of as a native capacity, an innate trait, or a function of general intelligence -- something we inherit, something we cannot change. Small wonder that talk of aptitudes among more egalitarian professionals in education has become increasingly rare. As evidence of how unfashionable aptitude had become, only a few

researchers have shown an interest in language aptitude during the twenty-year period between in 1970 and 1990, and the use of aptitude tests by foreign language teachers has significantly decreased (Ellis, 1994). It should be noted at this point that John B. Carroll, an expert on aptitude research in language testing, does not assume that aptitude is something innate,

...I am in general sympathy with writers like Neufeld (1978) who want to emphasize that foreign language aptitude, whatever it is, is not fixed or innate. They may be correct, and I would like to believe that they are. I am simply neutral on this matter, since we do not have the kind of evidence that would enable us to decide it... Yet, what evidence I have suggests that foreign language aptitude is relatively fixed over long periods of an individual's life span, and relatively hard to modify in any significant way. (Carroll 1997, p.86)

In the field of foreign and second language teaching, some of the disinterest in aptitude testing may have resulted from historical developments. The dominant models of foreign language aptitude were developed during the heyday of structuralist/behaviorist theories, and with their subsequent rejection a negative association with aptitude testing and research came about (Ellis, 1994). In the period before the 1970s, there were critics who maintained that factors such as motivation were responsible for differences in language learning performance (Hubbard, 1975; Corder, 1976; Roeming, 1966). Gardner and Lambert (1972) wrote the most memorable of these critiques,

One... wonders about the aptitude factor if he [sic] looks back into history a bit. When everyone had to know a second language, it seems that everyone, regardless of aptitude, learned it. Take France as an example. Many grandparents or even parents of today's generation spoke regional languages such as Basque, Breton, Provencal as home languages and learned French at school only, and apparently with no handicaps... It seems then that when the social setting demands it, people master a second language no matter what their aptitudes might be. What then is it to have a knack for learning a foreign language? (Gardner and Lambert, 1972, p.2)

It should be noted, however, that as persuasive as the foregoing argument may be, there is some hazard in conflating learning languages in a natural environment and learning languages in an academic environment (Cummins, 1983). Cummins (1983) made the distinction between basic interpersonal communication skill (BICS), where oral fluency and

socially-based interaction are the primary requirements, and cognitive academic language proficiency (CALP), where elements for success in an academic environment are dominant. It may be very difficult to identify aptitude in natural learning environments (Ellis, 1994; Carroll, 1979), such as the example of the regional French cited above; however, for language learning in academic environments there does exist a persuasive body of evidence that language learner performance can be predicted by a number of measures. The same R.C. Gardner in the Gardner and Lambert (1972) citation above, was party to a very different conclusion in studying cognitive variables in Gardner and MacIntyre (1992) concluding,

Research makes it clear that in the long run language aptitude is probably the single best predictor of achievement in a second language. (Gardner and MacIntyre 1992, p.215.)

## **To Stream or Not to Stream**

Embedded in any discussion of placement testing or aptitude testing is the ever present controversy of whether or not students should be tracked or streamed. After all, why would anyone employ these types of examinations unless students were being placed into ability streams or tracks? Later in this paper I will argue that ability streaming is not the only purpose for using placement and aptitude tests. However, before moving on to that alternative, I will first briefly summarize some of the debate about ability streaming.

From the development of statistical tools by Pearson (Grimm, 1993), through Galton's eugenics, to the development of the Stanford-Binet intelligence test by Terman, racism has been an ever present construct and streaming was one of its products (Oakes, 1986; Swanson, 1993). Ability streaming, for its many years of practice, does not have much support from the research community (Mosteller, Light and Sachs, 1996; Slavin, 1993; Swanson, 1993; Oakes, 1986). As offensive and misled as the history behind intelligence testing has been, it is surprising that assumptions regarding the innateness of intelligence and its capacity to predict learning performance have not come under closer inspection. In the case of learning second languages, is there an underlying assumption that "better" students are that way because they are more "intelligent?" One general argument against ability streaming is its potential for negative effects. Wesche (1979), echoing Vygotskian theory, notes that,

Grouping of individuals according to ability is controversial in almost any instructional situation. Ability grouping can improve the "fit" of the training

situation to the individual's capabilities at least with respect to pacing. Opponents of ability grouping point to the negative psychological effects of placement in slower groups, and cite the peer stimulation and help which can aid learning in a heterogeneous group, as well as the imperfections of existing ability measures as a basis of classification. (Wesche, 1979, p.123.)

In fact, early foreign language aptitude tests were little more than intelligence tests (Carroll, 1979). The Luria-Orleans Modern Language Prognosis Test (Luria and Orleans 1928) and the Foreign Language Prognosis Test (Symonds, 1930) were basically intelligence tests. These and other language aptitude tests fell out of use because they were ineffective at predicting language learning performance. It was not until the advent of factor analysis that the Modern Language Aptitude Test and the Pimsleur Language Aptitude Battery, tests focusing on language ability factors rather than on intelligence, were first developed and made available in the 1960s. While it must be admitted that for predicting language learning performance in an academic setting, there is some association between general intelligence and success in learning a foreign language (Skehan 1989, Carroll, 1979), conflating intelligence, even verbal intelligence, and foreign language aptitude is unjustified. Recent studies (Sparks, Ganschow, Patton, Atzer, Siebenhar and Plageman, 1997; Sparks, Ganschow and Patton, 1995, Service and Craik, 1993; Service, 1992) have found that factors such as phonology, short-term memory efficiency and similar factors are superior predictors of foreign language learning performance. These results suggest that testing which is heavily reliant on intelligence factors may not be assessing the most important factors which contribute to language learning performance.

The foregoing arguments support the contention that streaming predicated upon the inappropriate factors can potentially cause negative learning consequences. Three studies that will be mentioned in more detail later, have provided some evidence that streaming could be much more effectively employed when it is based upon learner needs than on factors such as intelligence or content knowledge (Sternberg, 1997; Wesche, 1979; Henrie, 1977.)

### **Prognostic or Diagnostic**

While the debate over ability streaming continues to rage, the possibility of grouping or streaming by other factors than "cognitive skill," such as by learning styles, has received very little attention. By learning styles I am referring to general ways that people differ in how they learn. There is a large body of literature addressing learning styles which is currently inconclusive, but points persuasively to significant differences in how people learn languages (Ellis, 1994).

One relevant analysis comes from Skehan (1989) who classified learners as analytic or memory-oriented. From his research, Skehan concluded that both types can successfully learn languages, but that they do so in fundamentally different ways. There is also evidence that learner performance will be enhanced by instruction which is in agreement with a subject's learning style and adversely affected from instruction which does not match learning style (Sternberg, 1997, Wesche, 1979;). Although it was a short term study unrelated to language learning, Sternberg (1997) found that when adult teaching students were tested on four unconventional placement factors (memory, analysis, creativity and practicality) the subjects who were matched to instruction had enhanced performance. Interestingly, it was found in Sternberg (1997) that two of the factors, creativity and practicality, improved the predictability of the testing on future performance. This raises the question of the construct validity of placement tests: have we been making placement decisions on the basis of the wrong factors?

The diagnostic potential in placement testing for a large second language learning program was examined by Wesche (1979). In a remarkable study, Wesche analyzed the 1976 data on nearly 3,000 students attending Canada's Public Service Commission (PSC) year-long intensive language training program for public service employees. Students were placed into four groups, each with a different teaching approach. Placement was based upon the Modern Language Aptitude Test (MLAT), two subtests from the Language Aptitude Battery (LAB), counselor interviews and a questionnaire. Data from these instruments were combined into a profile of each learner which was then used to match learners with a particular teaching approach. In an earlier study of the PSC program, Henrie (1977) found that a comparison between advanced learners in two of the groupings that students and teachers both reported a high level of satisfaction, and that there was no streaming effect -- there were no significant differences in achievement between the groups. Commenting on the Henrie (1977) study, Wesche concluded,

Since appropriate matching should lead to equal achievement by different types of individuals with comparable ability in different treatments, this was an encouraging finding. (Wesche 1979, p.137)

Wesche (1979) found that using aptitude test data helped to identify two special groups within the population which led to better matching of students to methodology. She also found that the aptitude testing improved the program by contributing better prognostic information as well as diagnostic information -- information that was used to improve instruction. Wesche (1979) also included a study of matched pairs (matched by aptitude profiles) in which one group was mismatched to the instructional methodology, resulting

in differences in student performance on achievement tests and in motivation,

Significant differences in the expected direction were found on almost all variables of interest. The appropriately matched students reported greater interest in foreign languages, more initiative to continue practicing French out of class, a more positive attitude toward the teaching method used, and less anxiety in class. They also achieved superior scores on three of the four achievement measures of listening comprehension and oral expression. (Wesche 1979, p137)

## Conclusion

The potential use of prognostic testing instruments for diagnostic purposes and subsequent applications for adjusted instruction was confirmed by two independent studies of the PSC, Henrie (1977) and Wesche (1979). Given the significance of these findings it does seem odd that interest in aptitude testing as a component of placement testing has not developed more interest. Given the results of these studies, the assertion that all we need to do in placement testing is to divide students into groups seems, to me, to be naively absurd. As we in the language teaching community continue to use testing opportunities to produce data of questionable value to place students, when we could be collecting data which potentially could transform our programs, we are incurring a great loss. The claim that the additional effort required for these testing and program measures is not justified, was clearly not the case in the PSC example. Given the development of teaching methods since the mid-1970s, we are in an even better position today to adapt instruction to the learning needs and learning styles of our students. Also, remarkable advances in test evaluation and item analysis makes it possible that even small language programs can efficiently develop better tests. As the PSC example and the streaming debate seem to suggest, should we choose not to improve our placement methods, less effective instructional intervention and unequal learner performance may very well be the result.

In this article I have been leading towards an advocacy of including additional factors into placement testing, specifically factors which will provide data on aptitude and learning style. Placement with these two additional factors hold great promise.



## References

- Bachman, L. F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Carroll, J. B. 1979. Twenty-five years of research on foreign language aptitude. In K. C. Diller (ed.) *Individual differences and universals in language learning aptitude*. Rowley, Mass.: Newbury House.
- Corder, S. P. 1967. The significance of learner's errors. In J. C. Richards (ed.) *Error analysis: perspectives on second language acquisition*. Essex: Longman.
- Cummins, J. 1983. Language proficiency and academic achievement. In J. Oller (ed.) *Issues in language testing research*. Rowley, Mass.: Newbury House.
- Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.
- Gardner, R. C. and Lambert, W. E. 1972. *Attitudes and motivation in second language learning*. Rowley, Mass.: Newbury House.
- Gardner, R. C. and MacIntyre. 1992. A student's contributions to second language learning. Part 1: Cognitive variables. *Language Teaching*, 25.
- Grimm, G. 1993. *Statistical applications for the behavioral sciences*. New York: John Wiley and Sons, Inc.
- Henrie, A. M. 1977. Rapport d'Evaluation: Approche Traditionnelle. Internal report, Evaluation Service, Public Service commission of Canada, Ottawa. [Reported on in Wesche, 1979].
- Hubbard, L. J. 1975. Aptitude, attitude, and sensitivity. *Foreign Language Annals*, 8, 33-37.
- Hughes, A. 1989. *Testing for the language teacher*. Cambridge: Cambridge University Press.
- Mosteller, F., Light, R., Sachs, J. 1996. Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review*, 66, 4, p797-842.
- Neufeld, G. G. 1978. A theoretical perspective on the nature of linguistic aptitude. *International Review of Applied Linguistics*, 16, 15-25.
- Oakes, J. 1986. *Keeping track: How schools structure inequality*. New Haven: Yale University Press.
- Roeming, R. F. 1966. The predictability of language learning results. *Contact*, 9 [ERIC Document Reproduction Service, ED 013 038]
- Service, E. 1992. Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology*, 45, 1, p.21-50.
- Service, E. and Craik, F. I. M. 1993. Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, 32, p.608-623.
- Skehan, P. 1989. *Individual differences in second-language learning*. London: Edward

Arnold.

- Slavin, R. E. 1993. Ability grouping in the middle grades: Achievement effects and alternatives. *Elementary School Journal*, 93, p.535-552.
- Sparks, R. L, Ganschow, L., Patton, J., Atzer, M., Siebenhar, D. and Plageman, M. 1997. Prediction of foreign language proficiency. *Journal of Educational Psychology*, 89, 3, p.549-561.
- Sparks, R. L, Ganschow, L., Patton, J. 1995. Prediction of Performance in first-year foreign language courses: connections between native and foreign language learning. *Journal of Educational Psychology*, 87, 4, p.638-655.
- Sternberg, R. J. 1997. What does it mean to be smart? *Educational Leadership*, 54, 6, p.20-24.
- Swanson, G. 1993. The hall of shame. *Phi Delta Kappan*, June 1993, p.796-798.
- Wesche, M. B. 1979. Language aptitude measures in streaming, matching student with methods, and diagnosis of learning problems. In K. C. Diller (ed.) *Individual differences and universals in language learning aptitude*. Rowley, Mass.: Newbury House.

[Received December 29, 1997]