# Centre for Canadian Language Benchmarks/
# Centre des niveaux de compétence linguistique canadiens

## An Examination of the Validity of the
## Canadian Language Benchmarks Placement Test

Executive Summary, August 14, 2001 by Phil Nagy,
The Ontario Institute for Studies in Education of the University of Toronto

An examination was undertaken of the levels of agreement between the Canadian Language Benchmark Placement Test (CLBPT) and the Canadian Language Benchmarks Assessment (CLBA), and the time saved by using the CLBPT.  The sample available, just over 500, was not as large as the 1600 that had been hoped for.  The sample was adequate to establish the overall level of agreement between the two instruments, but not the level of agreement at each benchmark.  We can be certain, with 95% confidence, that the CLBPT agrees perfectly with CLBA between 33% and 41% (37% ? 4%) of the time, and agrees within one benchmark between 79% and 85% (82% ?3%) of the time.  Considering that the CLBA is itself considered accurate within ? one benchmark, this is quite a good level of agreement.  Except that the CLBPT gave slightly higher average results in Listening/Speaking and in Reading at the middle benchmarks, the disagreements were symmetrically distributed. The CLBPT can be used for the same low stakes decisions for which the CLBA was designed.

Time saved with the shorter test increased with benchmark level.  Total testing time saved was from about 15 minutes at Benchmark 1 to 75 minutes at Benchmark 8.  However, there were numbers of cases where the CLBPT test took as long or longer than the CLBA test: for Listening/Speaking, 30% of respondents, for Reading, 20%, and for Writing, 20%.

Attempts at diagnosing the source of disagreement were hampered by the small samples.  Not only were there lower levels of test writing than expected, but the high level of agreement further lowered the available number of cases of disagreement.  Nevertheless, it was possible to identify some sites where Listening/Speaking agreement was lower than others.

Administrators at the pilot assessment sites provided substantial feedback on the validation process.  Many of the difficulties identified focused on the lack of time available for the validation.  The tapes used in the Listening/Speaking tests, and the written forms for all tests, need to be upgraded to production quality.  Test instructions and training procedures require clarification and fine tuning.  Some test items require revisions.  A working group has been struck to deal with these revisions as prelude to the implementation process.  Further validation data will be collected as required.