DOCUMENT RESUME

ED 476 424 TM 034 926

AUTHOR	Emenogu, Barnabas; Childs, Ruth A.
TITLE	Curriculum and Translation Differential Item Functioning: A Comparison of Two DIF Detection Techniques.
PUB DATE	2003-04-00
NOTE	39p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 22-24, 2003).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Curriculum; English; Foreign Countries; French; French Canadians; *Item Bias; Item Response Theory; Mathematics Tests; Standardized Tests; State Programs; *Test Format; Testing Programs; *Translation
IDENTIFIERS	Mantel Haenszel Procedure; Ontario

ABSTRACT

This study investigated the possible impacts of language and curriculum differences on the performance of test items by subpopulations of students. Focusing on Measurement and Geometry items completed by students in French- and English-language schools in Ontario made it possible to explore the differences and to compare the item response theory (IRT) and Mantel-Haenszel (MH) approaches to finding items with differential item functioning (DIF). The tests in this study might have had DIF as a result of translation or curriculum. Data came from the 2001 administration of the School Achievement Indicators Program (SAIP) mathematics assessment. The dataset contained information on 793 13-year-old and 677 16-year-old students from English-language schools and 487 13-year-old and 546 16-year-old students from French-language schools. Thirteen of the 27 items studies were flagged as exhibiting DIF by the MH approach, and 6 of these 13 were also flagged by IRT. Differences in curricula and practice between the French- and English-language schools are very difficult to separate from language differences, and, in fact, results point to the possibility of another source of DIF: different test-taking approaches in the two populations. Results demonstrate the complexity of the factors that contribute to how items are understood and approached by different groups of students. (Contains 5 tables, 7 figures, and 40 references.) (SLD)

Running Head: Curriculum and Translation DIF

Curriculum and Translation Differential Item Functioning:

A Comparison of Two DIF Detection Techniques

Barnabas Emenogu and Ruth A. Childs

Ontario Institute for Studies in Education of the University of Toronto

23 April 2003

A Paper presented at the 2003 annual meeting of the National Council on Measurement in Education,

Chicago, IL, April 21-24, 2003

Correspondence concerning this article should be addressed to Ruth A. Childs, OISE/UT,

252 Bloor Street West, 11th Floor, Toronto, Ontario M5S 1V6. E-mail: rchilds@oise.utoronto.ca

Curriculum and Translation Differential Item Functioning:

A Comparison of Two DIF Detection Techniques

Introduction

It is usual to worry about translation-related differential item functioning (DIF) when tests are given in different languages. Indeed, previous studies have stressed the importance of identifying sources of translation DIF (Allalouf, 2000; Allalouf, Hambleton, & Sireci, 1999; Gierl, & Khaliq, 2001; Gierl, Rogers, & Klinger, 1999). However, translation problems may not be the right explanation for all items that exhibit DIF.

Many studies (e.g., Huang, 1998; Mehrens & Phillips, 1986; Miller & Linn, 1988) suggest that the degree of match between an assessment and the taught curriculum can have a large impact on achievement test scores. Huang (1998), for instance, showed that controlling the error variance due to curriculum sampling decreased slightly the rate of item classification inconsistency, and asserted that the "finding suggested that different school curriculum may play a role in the differences found in student performance" (p.13). Porter, Schmidt, Floden and Freeman (1978) assert that state or provincial decision makers need to be keenly aware of the relationship between tests and curricula, in addition to translation issues, if they are to make sense of test results from schools using different languages. This is especially so where instruction occurs in different languages and curriculum-aligned materials are differentially available across language groups. Although some items may be translated very accurately, they may still behave differently because of differences in how the curriculum is defined and/or taught in the different language groups. The School Achievement Indicators Program (SAIP) mathematics test results from Ontario provide a unique opportunity to study translation and curriculum DIF, because we know something about differences in curricula between French- and English-language schools.

There does not seem to be agreement among psychometric professionals on the statistical procedures best suited for DIF detection. Psychometric literature is rich with studies that use one approach or another or, in some cases, more than one approach to investigate DIF (Narayan & Swaminathan, 1996; Prieto, Barbero, & San Luis, 1997; Raju, Drasgow, & Slinde, 1993; Shepard, Camilli, & Averill, 1981; Whitmore & Schumacker, 1999). The results sometimes point in different directions. However, Prieto, Barbero, and San Luis, (1997) observed that the most promising approaches appear to be the Mantel-Haenszel (M-H) procedure and those based on the fundamentals of item response theory (IRT). The IRT methods are theoretically preferred but are computationally intensive and require a minimum sample size of 1000 examinees and test length of 40 items (Shepard, Camilli & Averill 1981; Raju 1988, 1990). On the other hand, the M-H approach has also become popular because it is easy to implement and has an associated test of statistical significance (Prieto, Barbero, & San Luis, 1997). While Hambleton and Rogers, (1989), Swaminathan and Rogers (1990) and Narayanan and Swaminathan (1996) showed that the M-H procedure was not very effective in identifying nonuniform DIF items, Prieto, Barbero, and San Luis (1997) found that the M-H procedure may be effective for detecting a relatively high proportion of nonuniform DIF. In an empirical comparison study, Raju, Drasgow, and Slinde (1993) found that there was close agreement between IRT-based procedures and the M-H procedure in a female-male comparison while different items were identified for DIF by the two procedures in their black-white comparison. McLaughlin and Drasgow (1987) and Lim and Drasgow (1990) showed that when Lord's chi-square statistic is employed with joint maximum likelihood estimates of ability parameters, it leads to incorrect and misleading DIF results, but when used with marginal maximum likelihood or Bayes modal estimates of ability parameters, it yields more nearly accurate DIF results. The point being made is that the search for effective methods for determining DIF items is inconclusive. Rogers' (1989) proposal for further studies to be conducted with these methods when the focal and reference groups have unequal ability, when sample size is small or when test length is short, is still relevant.

Objectives of the Study

The objectives of this study, therefore, are;

1. to use both item response theory (IRT) and Mantel-Haenszel (MH) approaches to look for DIF items,
2. to compare the numbers and patterns of items identified by the two approaches,

3. to examine the degree to which these patterns are related to the differences we expect to find given what we know about differences in the curricula, and
4. to explore what may be at the root of DIF in the large-scale examination, and how significant results may be interpreted.

Educational Importance of the Study

Muthén, Kao, and Burstein (1991) have called attention to instructional sensitivity of test items due to opportunity to learn. Citing Linn and Harnish (1981), they warned against mistakenly attributing DIF due to instructional bias to other sources such as ethnicity. As well, an earlier study by Shen (2000) revealed the influence of curriculum type on differential item functioning. However, as observed by Sireci and Swaminathan (1996), there are two problems to contend with when conducting a DIF study with two different language groups of examinees on different language versions of test items. Here one must try to separate the effects of curriculum differences and item language differences.

The unique contribution of the present study lies in the fact that it compares the IRT and MH approaches to helping untangle DIF in tests that may have both translation and curriculum DIF. It further reinforces the idea that possible sources of DIF can be identified using a general IRT loglinear regression as well as the Mantel-Haenszel approaches and compares the results of the two approaches.

Method

Data

This study uses data from the content subtest of the 2001 SAIP Mathematics Assessment. Students who wrote the content subtest received a booklet containing 27 background questions, 15 multiple-choice placement items and 110 short answer and multiple-choice items grouped into five sections according to difficulty. The items are ordered by level of difficulty. Of the 125 items on the test, 75 were multiple-choice items with four response options and 50 were short answer items.

Students were first administered the placement items. An exam proctor immediately scored these items. Based on the score on the placement items, each student was told to continue the test from one of three "starting points" – Item 16, Item 41, or Item 66 – and to work as far as possible within the test time (CMEC, 2001).

The content subtest assesses student achievement in four areas: (1) Numbers and Operations, (2) Algebra and Functions, (3) Measurement and Geometry, and (4) Data Management and Statistics. The Measurement and Geometry (M&G) items are the focus of this study. The content subtest has a total of 31 M&G items; however, four of these are in the placement test. This study examines the 27 M&G items that follow the placement test.

Sample

For the administration of the 2001 SAIP Mathematics Assessment, students were sampled from each participating province and, within some provinces, by language of instruction. Each student took either the content or the problem solving subtests, but not both.

To facilitate the examination of DIF caused by curriculum and language differences, it was important to define two relatively homogeneous groups for comparison. The Ontario students in English-language and in French-language schools were selected.

Of the Ontario students who took the test, those who omitted all 15 of the placement test items or did not provide their age were excluded. The resulting dataset consisted of 793 13-year-old and 677 16-year-old students from English-language schools and 487 13-year-old and 546 16-year-old students from French-language schools. Because only about 5% of Ontario students are enrolled in French-language schools, in the 2001 SAIP data collection, the students from Ontario's French-language schools were deliberately over-sampled to provide a number large enough for group analyses. Consequently, a student in a French-language school was almost 15 times as likely as a student in an English-language school to participate in the 2001 SAIP Mathematics Assessment.

For the analyses, in addition to dividing students by age and language, it was necessary to divide the students by starting point in order to minimise the missing data in each analysis. The placement test items were not included in the analyses because performance on those items showed very little variability, particularly for students assigned to begin at Starting Points 2 (Item 41) and 3 (Item 66). For the students beginning at each starting point, only the M&G items falling within the 60 items following that starting point were analysed, as few students were able to work farther than that within the time allowed.

DIF Analyses

Mantel-Haenszel (M-H) approach. An M-H chi-square (Holland & Thayer, 1988) was computed for each item, using the EZDIF program (Waller, 1998). The M-H computations involve testing whether the odds of answering an item correctly at a given score level is independent of group membership, when groups are matched on ability or achievement. This involves obtaining an odds ratio, α_i,for each item in both the focal and the reference groups.

$$\alpha_i = \frac{p_{ri} q_{fi}}{p_{fi} q_{ri}} \tag{1}$$

where

p_{ri} and q_{ri} represent the proportions in the focal group that responded correctly and incorrectly to the *i*th item, and

p_{fi} and q_{fi} represent the proportions in the reference group that responded correctly and incorrectly to the *i*th item.

A chi-square distribution with df=1 and α=.01 was used to identify items with significant DIF.

Item Response Theory (IRT) approach. IRT provides a class of models that describe the relationship between latent ability and the probability of correctly answering an item. This probability is determined by factors referred to as item parameters, which can include all three or only one or two of *item discrimination, item difficulty* and a *guessing factor*. In the most general case where all three parameters apply, the model is the *three parameter logistic* (3PL) model, which is expressed as

$$P_j(\theta) = c_j + \frac{1 - c_j}{1 + e^{-Da_j(\theta - b_j)}} , \tag{2}$$

where

c_j is the probability of getting the item right just by guessing,

b_j, the intercept of the trace line, is the difficulty parameter,

a_j or the slope of the trace line is the discrimination parameter, and

θ is the latent ability or proficiency.

When c_j is fixed to zero, the 3PL model simplifies to the 2PL model.

A number of DIF detection procedures are classified as IRT methods. Among them are Lord's chi-square statistic, the signed area measure and the unsigned area measure (Raju, Draslow, & Slinde, 1993), as well as the model comparison approach described by Thissen, Steinberg and Wainer (1993). Marginal Maximum Likelihood (MML) item parameter estimates were computed using MULTILOG (Thissen, 1991). As described above, the content sub-test includes both multiple-choice and short answer items. The 3PL model was considered appropriate for the multiple-choice items; the 2PL model, for the short answer items. Omitted items within the 60 items following a student's assigned starting point were counted as wrong.

To examine items for DIF, the process described by Thissen, Steinberg and Wainer (1993) for comparison of two models was followed. An *augmented model* (A) includes all the parameters of a *compact* model (C) plus additional parameters. The analysis tests whether the additional parameters in the augmented model result in a significant improvement in model fit. The test of significance is of the form

$$G^2(df) = 2 \log \left[\frac{Likelihood[A]}{Likelihood[C]} \right], \tag{3}$$

where *Likelihood*[.] represents the likelihood of the data given the maximum likelihood estimates of the model and *df* is the difference between the number of parameters in the augmented model and the number of parameters in the compact model. Under very general assumptions, the value of $G^2(df)$ is distributed as $\chi^2(df)$ under the null hypothesis. Thus, if the value of $G^2(df)$ is large, representing an unlikely value from a $\chi^2(df)$ distribution, we reject the null hypothesis and the compact model (Thissen, Steinberg, & Wainer, 1993).

To equate the parameters of both the focal and reference groups in the same metric, the *anchor test method* (Camilli & Shepard, 1994), whereby item parameters for both groups are simultaneously estimated, was used. This approach requires constraining the parameters of the *anchor items* to be identical for the two groups, while the parameters of one item, the *studied item,* are allowed to vary. In these analyses, the item parameter estimates for the English- and French-language versions of the items were constrained to be equal for all items except the particular item being studied. In addition, the item parameters were estimated with all items constrained to be equal across the French- and English-language versions. This was the compact model. The model in which one item's parameters were allowed to vary was the augmented model. If the fit of the augmented model was significantly better than that of the compact model – that is, allowing separate parameters for the two versions of the item resulted in a significantly improved fit – the item was considered to exhibit DIF.

Results and Discussion

Items Exhibiting DIF

Thirteen of the 27 Measurement and Geometry items were flagged as exhibiting DIF by the Mantel-Haenszel approach (see Table 1). Six of these 13 items were also flagged by the IRT approach. No additional items were flagged by the IRT approach. One of the items (Item 33) was flagged for both 13- and 16-year-old students assigned to Starting Point 1. No other items were flagged for both ages or for multiple starting points.

Translation

Each item on the SAIP was developed in either English or French and was then translated into the other language. As the report of the 1997 mathematics assessment (CMEC, 1997) describes,

> A linguistic analysis of each question and problem was also conducted to make sure French and English items functioned in the same manner. For the marking sessions, francophone and anglophone coders were jointly trained and did the marking together in teams working in the same rooms. (p. 4)

We would expect these efforts to minimise the possible sources of translation DIF. As other studies (e.g., Allalouf, Hambleton, & Sireci, 1999; Gierl & Khaliq, 2001) has shown, however, it is very difficult to achieve perfect agreement in the meaning and vocabulary difficulty of translated materials.

One challenge for students in French-language schools across Canada is differences in vocabulary use between provinces. Because the French-speaking communities outside of Quebec tend to be small and, often, isolated, the evolution of French words in these communities does not always parallel that in Quebec. The French-language versions of most SAIP mathematics items were created and approved by individuals speaking French as it is spoken in Quebec. It is possible that some of the vocabulary used was less familiar and therefore more difficult for Ontario students in French-language schools than for Quebec students. Furthermore, if the focus was implicitly on ensuring that items were of the same difficulty for students in English-language schools outside Quebec and students in French-language schools in Quebec, the students in French-language schools in Ontario may well be disadvantaged in comparison to students in English-language schools in Ontario. This would suggest, however, that the items exhibiting DIF should favour students in English-language schools. In fact, seven of the 13 items favoured the students in French-language schools.

Beyond vocabulary difficulty, there may be more subtle translation differences. A review of the items suggests some possible differences. For example, Item 35 requires identifying "lines" or "lignes" in a drawing. Some of the lines are curved and it may be that students taking the French-language version were more likely to assume that only straight lines were to be counted as lines. Item 49 involves edges, faces, and vertices of a three-dimensional object. This item was significantly more difficult for the 16-year-old students taking the English-language version of the test. An examination of the responses indicates that 10% of the students taking the test in French confused faces and vertices in their answer, while more than 20% of those taking it in English made this mistake. It seems that the term "side" is often used instead of "face" in the materials for the English-language students, which may have contributed to the confusion.

Curriculum

In 1999, Ontario introduced a new high school curriculum. The curriculum was introduced for that year's Grade 9 students, but did not apply to earlier cohorts of students – that is, students in Grade 10 or higher in 1999. According to Ontario's "context statement", at the time of the 2001 SAIP Mathematics Assessment

> "...most 13-year-old students were enrolled in either grade 8 or grade 9 mathematics both of which are mandatory core subjects in the new curriculum ... [however] most of the 16-year-old students in the assessment would have been studying the old mathematics curriculum and taking a grade 11 course at one of the three possible levels of difficulty or would have taken no mathematics course since grade 10" (CMEC, 2001, p. 57).

The new curriculum differs from the old both in its content and in the process used to develop it. Before 1997, the provincial mathematics curriculum in Ontario was developed in English, then translated into French (Ontario Ministry of Education, 1985a, 1985b). As a result, in both languages, the defined curriculum contained the same content, although differences existed in how content was presented in textbooks and other resource materials, and different resources were available in English and in French. The post-1997 mathematics curriculum (Ontario Ministry of Education, 1997a, 1997b, 1999a, 1999b, 2000a, 2000b), in contrast, was developed separately for French-language and English-language schools. The curriculum development teams worked in parallel and most of the expectations are for the same content. However, a few expectations differ. The example that follows illustrates the difference. The English-language curriculum explicitly states that both Grade 9 Academic and Applied courses will expect students to "substitute into and evaluate algebraic expressions involving exponents, to support other topics of the course (e.g., measurement, analytic geometry)". But the French language curriculum does not have this expectation (although this content might be implied from some of the other expectations). Additionally, in French language Grade 9 Applied courses, students must "communiquer les étapes de la résolution de problèmes et les justifier" and "vérifier la solution d'une équation," while the same expectations are not explicit for French Academic or English Applied or Academic courses. As these examples illustrate, the differences are not large, but may well result in different content being taught and with different emphasis. It may, in fact, be that the differences in curricula reflect differences that already existed when the curricula were developed.

The materials to which classroom teachers have access also differ. Many more textbooks are available in English in Ontario. However, French-language school boards have often collaborated in developing curriculum-linked materials for classroom teachers. Because fewer textbooks and other resources are available to the teachers in French-language schools, the board-developed materials are very detailed. Because little else is available, the teachers tend to rely on these materials. As a result, teachers of mathematics in French-language schools are more consistent across the province in the content and style of the instruction they deliver than are teachers in English-language schools.

Differences in curricula and practice between French- and English-language schools are very difficult to separate from language differences. To further complicate things, it is not clear to what extent teachers in either language are teaching the new curriculum. In fact, some of the SAIP items are not covered explicitly in the new curriculum (Ontario Ministry of Education, 2002) and so might favour students still being taught under the old curriculum. For example, Item 49, which involves edges, faces, and vertices of a three-dimensional object, presents a problem in that the term "edges" is not used in the new curriculum. Item 84 requires solving a problem relating the dimensions of a cylinder to its volume – this content is in the new Grade 9 curriculum, so that the 13-year-old students taking the test may not have been familiar with it. Item 101 involves the lengths and angles between chords of a circle, content that was included in neither Ontario's old nor new curriculum. Similar mismatches between item content and curriculum were identified by Schmidt, Wolfe, and Kifer (1992) on the Second International Mathematics Study (SIMS) and by Lawson, Bordignon, and Nagy (2002) on the Third International Mathematics and Science Study (TIMSS).

Differences in item functioning may also have to do with what students know, but in areas not directly related to the curriculum. Item 35, for example, had as its graphic a map of the Quebec City area and was slightly easier for students taking the French-language version.

Test-Taking Approach

This study set out to examine translation and curriculum as sources of DIF in the SAIP Mathematics Assessment. However, the analyses suggest an additional possible source of DIF for these items: students' test-taking approaches.

As Figure 1 shows, students beginning at Starting Point 2 and taking the English-language version of the test attempted more items than students taking the French-language version did. This was true for both 13- and 16-year-old students and for all three starting points. This makes interpretation of the results difficult. According to Table 1, the items later in the test tended to favour students taking the English-language version; however, more students taking that version attempted these items. In fact, the percentages of students answering each item correctly out of those who attempted it, are higher for students taking the French-language version, indicating that they were more likely, if they responded to an item, to respond correctly. Different explanations are possible. It may be that greater difficulty of the vocabulary for the Ontario students taking the French-language version resulted in slower responding. However, it is also possible that other factors, such as experience with similar tests or a lesser propensity to guess, contributed to a different test-taking approach.

H. Jodouin, who is an on-site observer of the administration of the SAIP assessments in both English-language and French-language schools, notes that students in the English-language version tend to be encouraged to attempt as many items as possible and to give their best guess on the multiple-choice items. According to him, administrators of the French-language version tend to remind students that they should be able to justify their answers (H. Jodouin, personal communication, 13 January 2003). Other Ontario educators have suggested that students taking the French-language version may work more slowly on items because their teachers and textbooks emphasise careful and thorough work.

Results from the Second International Mathematics Study (SIMS) revealed large differences in the number of items attempted across countries (Schmidt, Wolfe, & Kifer, 1992). For example, students in France showed the highest omit rates, with almost half of the students omitting some of the items; students in Thailand, in contrast, were very unlikely to omit, with most items having fewer than one percent omits. As Schmidt, Wolfe, and Kifer (1992) note, such differences can have a significant impact on test results, particularly for multiple-choice items. They describe the possible effect of differences in test-taking approach as follows:

> Students in one system may, for a number of reasons, be willing to answer a multiple-choice question on the mathematics test even if they are not sure they really know the answer. Because of cultural differences, students in another system with the same level of knowledge and certainty might choose to not answer the question. In the first case, even if they do not know the answer, they have a non-zero probability of getting the item correct; such is the role of chance in multiple-choice testing. The students in the second culture have a zero probability of getting the item correct even if they know the answer. Hence, answering and knowing are not the same thing! (p. 88)

Similarities and Differences between IRT and MH Procedures

Both IRT and MH procedures gave exactly the same result for only two items, Items 33 and 35 for Starting Point 1 for the 13-year-olds. Item 33 showed DIF at the .001 level of significance in both IRT and MH procedures in favour of students in English-Language schools. Item 35 showed DIF only at the .05 level of significance in both procedures in favour of students in French-Language schools. Item 33 is a short answer, conceptual question, while Item 35 is a short answer procedural question. Based on their *b* parameters, both are not very difficult items and have somewhat low levels of discrimination, ranging from 0.37 for French Item 35 to 0.76 for English Item 33. In terms of the magnitude of DIF, Item 33 is the one that really gives cause for concern, being flagged at the .001 level of significant by both procedures. There are no other items flagged at the .001 level of significance by both procedures.

Four other items, 105 (for the 13-year-olds at Starting Point 3), 49 and 94 (for the 16-year-olds at Start Point 2), and 101 (for the 16-year-olds at Start Point 3) were flagged at the .01 level of significance by the MH procedure. Incidentally, these same items were flagged at the .05 level of significance by the IRT procedure, and all except Item 49 favoured students in English-Language schools. In summary, the MH approach identified more items (5 items) with DIF at the .01 level or less, while the IRT procedure identified only one at that level. Of the five items, two are short answer and three are multiple choice. All but one measures conceptual ability; the exception measures procedural ability.

It is important to note that those four items flagged by the MH procedure at the .01 level of significance were also flagged by the IRT procedure at the .05 level of significance. This suggests that we have a higher chance of not rejecting a false null hypothesis, thereby committing a Type II error, by using the IRT approach than by using the MH procedure. In a previous study Raju, Drasgow and Slinde (1993) compared MH and three IRT approaches in detecting race-based (Black-White) and gender-based (Female-Male) DIF. In the Black-White comparison, they found that the MH approach identified two items at the .001 level of significance while the three IRT approaches identified only one item at the .001 level of significance. There was no overlap of biased items between any of the IRT approaches and the MH approach, while there was 100% overlap between all three IRT approaches. In the Female-Male comparison, whereas the MH approach also identified one more DIF item than the IRT approaches, there was a 100% overlap of biased items between the IRT approaches and the MH approach.

There is no clear pattern observed in the identification of DIF items by IRT and MH procedures in relation to their relative chi-square values. For instance, one of the five items identified by MH at the .01 level of significance and by IRT at the .05 level of significance has a smaller MH chi-square value ($\chi^2_{M\text{-}H}$) than IRT chi-square value (χ^2_{diff}). Also, of the 8 items flagged by MH at the .05 level of significance but not flagged by IRT, 3 have higher $\chi^2_{M\text{-}H}$ values than χ^2_{diff} values, while 5 have higher χ^2_{diff} values than $\chi^2_{M\text{-}H}$ values.

Hills (1989), Raju (1990) and Shepard, Camilli and Averill (1981) have suggested that the IRT approaches are theoretically preferred because of the invariance property of IRT item parameters. However, Hills (1989) and Prieto, Barbaro and San Luis (1997) are of the view that the MH techniques are more practical because they are easy to implement, have an associated test of significance and are relatively stable with small sample sizes and short test lengths. The Prieto, Barbaro and San Luis (1997) study also found, in contrast to Swaminathan and Rogers (1990), that the MH procedures can be effective in detecting a relatively high proportion of nonuniform DIF.

Limitations

This study has several limitations. First, in order to obtain sufficiently complete data matrices, the analyses were performed by age and by starting point. This meant, however, that the numbers of students in some groups were quite small, likely impacting the stability of the DIF analysis results. Second, the decision to treat omitted items as wrong made certain assumptions about the test-taking approaches of the students taking the test. The dramatically different numbers of students attempting the later items on the French-language and the English-language versions of the test suggest that these two groups may have been using different test-taking strategies. This possibility merits further investigation. Finally, the correspondence between the prescribed curriculum and what is actually taught in the classroom is rarely perfect. Although few differences were found between mathematics curricula for the two languages of instruction in Ontario, it is not easy to dismiss curriculum as a possible source of DIF without analysing actual classroom experiences of the two groups.

Conclusion

The purpose of this study was to investigate the possible impacts not only of language, but also of curriculum differences, on the performance of test items for subpopulations of students. By focusing on Measurement and Geometry items and students in French- and English-language schools in Ontario, we were able to explore such differences more closely than has been possible for studies using more heterogeneous sets of items and samples of students. The results demonstrate the complexity of the factors that contribute to how items are understood and approached by different groups of students.

The results also suggest areas for further exploration. First, curricular differences might be better understood in combination with information about teachers' classroom practices. Teachers' academic training, experience, and the materials available to them might well influence practices. Such contextual information would help us understand how the curriculum is being understood and presented. In addition, an examination of the patterns of items attempted by students taking the English- and French-language versions suggests a difference in test-taking approaches. Further research on the test-taking approaches of these two groups might well explain some of the differences in test results.

Author Note

The authors would like to thank the Council of Ministers of Education, Canada for providing the data. Also, thanks are due to Alana Cote, Hervé Jodouin, Kelsey Saunders, and Jacques Theoret of Ontario's Education Quality and Accountability Office and to Gilles Fournier of the CMEC for their helpful suggestions throughout this study.

Correspondence concerning this article should be addressed to Ruth A. Childs, OISE/UT, 252 Bloor Street West, 11th Floor, Toronto, Ontario M5S 1V6. E-mail: rchilds@oise.utoronto.ca

References

Allalouf, A. (2000, April). *Retaining translated verbal reasoning items by revising DIF items.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Allalouf, A., Hambleton, R., & Sireci, S. (1999). Identifying the causes of DIF on verbal items. *Journal of Educational Measurement, 36*, 185-198.

Council of Ministers of Education, Canada. (1997). *School Achievement Indicators Program (SAIP) 1997 Mathematics II Assessment.* Toronto, ON: Author.

Council of Ministers of Education, Canada. (2001). *Report on Mathematics Assessment III. School Achievement Indicators Program (SAIP) 2001.* Toronto, ON: Author.

Gierl, M. J., & Khaliq, S. N. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests: A confirmatory analysis. *Journal of Educational Measurement, 38*, 164-187.

Gierl, M. J., Rogers, W. T., & Klinger, D. A. (1999). Using statistical and judgmental reviews to identify and interpret differential item functioning. *The Alberta Journal of Educational Research, 45*, 353-76.

Hambleton, R. K. & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. *Applied Measurement in Education, 2*, 313-34.

Hills, J. R. (1989). Screening for potentially biased items in testing programs. *Educational Measurement: Issues and Practice, 8*(4), 5-11.

Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.

Huang, C. (1998, April). *Factors influencing the reliability of DIF detection methods.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Lawson, A., Bordignon, C., & Nagy, P. (2002). Matching the Grade 8 TIMSS item pool to the Ontario curriculum. *Studies in Educational Evaluation, 28,* 87-102.

Lim, R. G. & Drasgow, F. (1990). Evaluation of two methods of estimating item response theory parameters when assessing differential item functioning. *Journal of Applied Psychology, 2*, 164-174.

McLaughlin, M. E. & Drasgow, F. (1987). Lord's chi-square test of item bias with estimated and with known person parameters. *Applied Psychological Measurement, 11*, 161-173.

Mehrens, W. A., & Phillips, S. E. (1986). Detecting impacts of curricular differences in achievement test data. *Journal of Educational Measurement, 23*, 185-196.

Miller, M. D., & Linn, R. L. (1988). Invariance of item characteristic functions with variations in instructional coverage. *Journal of Educational Measurement, 25*, 205-219.

Muthén, B. O., Kao, C., & Burstein, L. (1991). Instructionally sensitive psychometrics: Applications of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement, 28*, 1-22.

Narayanan, P., & Swaminathan, H. (1996). Identification of items that show DIF. *Applied Psychological Measurement, 20*, 257-74.

Ontario Ministry of Education. (1985a). *Curriculum guideline, mathematics intermediate and senior divisions: Part two, Grades 7and 8, Grades 9 and 10, general level, Grades 11 and 12, general level.* Toronto, ON: Author.

Ontario Ministry of Education. (1985b). *Programme-cadre mathématiques cycles intermédiaire et supérieur: Deuxième partie, 7ᵉ et 8ᵉ année, 9ᵉ et 10ᵉ année, niveau général, 11ᵉ et 12ᵉ année, niveau général.* Toronto, ON: Author.

Ontario Ministry of Education. (1997a). *Le curriculum de l'Ontario, de la 1re a la 8e année – Mathématiques.* Toronto, ON: Author.

Ontario Ministry of Education. (1997b). *The Ontario curriculum, Grades 1-8 – Mathematics.* Toronto, ON: Author.

Ontario Ministry of Education. (1999a). *Le curriculum de l'Ontario, 9ᵉ et 10ᵉ année – Mathématiques.* Toronto, ON: Author.

Ontario Ministry of Education. (1999b). *The Ontario curriculum, Grades 9 and 10 – Mathematics.* Toronto, ON: Author.

Ontario Ministry of Education. (2000a). *Le curriculum de l'Ontario, 11ᵉ et 12ᵉ année – Mathématiques.* Toronto, ON: Author.

Ontario Ministry of Education. (2000b). *The Ontario curriculum, Grades 11 and 12, Mathematics.* Toronto, ON: Author.

Ontario Ministry of Education. (2002). *SAIP 2001 alignment with Ontario's mathematics curriculum.* Toronto, ON: Author.

Porter, A. C., Schmidt, W. H., Floden, R. E., & Freeman, D. J. (1978). *Impact on what? The importance of content covered* (Research Series No.2). East Lansing, MI: Michigan State University, Institute for Research on Teaching.

Price, L. R. (1999, April). *Differential functioning of items and tests versus the Mantel-Haenszel technique for detecting differential item functioning in translated test.* Paper presented at the annual meeting of the Alliance of Health, Physical Education, Recreation, and Dance, Boston, MA.

Prieto, P., Barbero, M. I., & San Luis, C. (1997). Identification of nonuniform differential item functioning: A comparison of Mantel-Haenszel and item response theory analysis procedures. *Educational and Psychological Measurement, 57*, 559-568.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207.

Raju, N. S., Drasgow, F., & Slinde, J. A. (1990). An empirical comparison of the area methods, Lord's chi-square test, and the Mantel-Haenszel technique for assessing differential item functioning. *Educational and Psychological Measurement, 53*, 301-314.

Rogers, H. J. (1989). A logistic regression procedure for detecting item bias. *Dessertation Abstracts International, 50*, 3928A. (University Microfilms No. 90-11; 788).

Schmidt, W. H., Wolfe, R. G., & Kifer, E. (1992). The identification and description of student growth in mathematics achievement. In L. Burstein (Ed.), *The IEA Study of Mathematics III: Student growth and classroom processes* (pp. 59-99). Oxford, England: Pergamon.

Shen, L. (2000, April). *Curriculum types as a differentiating factor in medical licensing examinations.* Paper presented at the annual conference of American educational Research Association, New Orleans, LA.

Sireci, S. G., & Swaminathan, H. (1996, October). *Evaluating translation equivalence: So what's the big dif?* Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.

Swaminathan, H., & Rojers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370

Thissen, D. (1991). *MULTILOG-6.0* [Computer software]. Chicago, IL: Scientific Software International.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale, NJ: Erlbaum.

Waller, N. G. (1998). EZDIF: Detection of uniform and non-uniform differential item functioning with the Mantel-Haenszel and logistic regression procedures [Computer Program Exchange]. *Applied Psychological Measurement, 22,* 391.

Whitmore, M. L., Schumacker, R. E. (1999). A comparison of logistic regression and analysis of variance differential item functioning detection methods. *Educational and Psychological Measurement, 59,* 910-927.

Table 1

The Measurement and Geometry Assessment Domain of the 2001 SAIP Mathematics Assessment

Criteria

Level 1

The student...

1. estimates and measures lengths and areas in terms of non-conventional units.
2. estimates and measures lengths in metres, decimetres or centimetres.
3. demonstrates an understanding of the concept of magnitude and the concept of area.
4. demonstrates an understanding of the concepts of interior, exterior and boundary.
5. with the aid of concrete materials, derives certain characteristics such as colour, form and size for solid forms and for two-dimensional figures.
6. demonstrates an understanding of the concept of symmetry in simple activities such as folding and the completion of drawings.

Level 2

The student...

1. estimates and measures dimensions of familiar objects, using SI units.
2. establishes relationships between SI length units.
3. demonstrates an understanding of the notion of volume.
4. estimates and calculates areas in both conventional and non-conventional units.
5. uses SI symbols, including the prefixes milli, centi, deci and kilo, in the context of length and area measurements.
6. solves simple real-life problems using SI units of length, area and time.
7. demonstrates an understanding of the concepts of angle and measures of an angle.
8. develops and applies problem-solving strategies related to spatial relations.
9. knows and uses characteristics of various solid forms for the purpose of classifying the solid forms.
10. determines, starting from concrete materials, certain characteristics of solid forms and the two-dimensional figures that form their boundaries.
11. constructs solid forms from simpler forms.
12. describes, draws, and classifies polygons and polyhedra according to certain of their properties.
13. describes and performs single geometric transformations of translation, rotation or reflection.

Criteria

Level 3

The student...

1. using concrete examples, demonstrates and understanding of the concept of the measure of an anle.
2. solves real-life problems relating to angle measure, segment length, area, volume, mass and time.
3. reproduces plane figures by using repetition of one of the plane geometric transformations of translation, rotation, reflection or similarity.
4. constructs various plane geometric figures.
5. identifies and uses various characteristics relating to triangles, quadrilaterals and circles.
6. calculates area, circumference and diameter of a circle.
7. establishes relationships between SI units (including squared and cubic units).

Level 4

The student...

1. solves problems derived from real-life situations that use the characteristics of solid forms.
2. finds the unknown elements in right-angled triangles and isosceles triangles using scale drawing or calculation methods, provided that the student is given the option of which method he/she can use.
3. solves real-life problems using the notion and units of capacity.
4. solves problems derived from real-life situations involving area, circumference and diameter of a circle.
5. reproduces plane figures using ordered sequences of the following plane transformations: translation, reflection, rotation and similarity
6. demonstrates an understanding of the concepts of isometry (congruence) and similitude (similarity).
7. solves real-life problems using the concepts of isometry and similitude in triangles and polygons.

Level 5

The student...

1. calculates arc and chord lengths, sector and segment areas, associated with a circle.
2. solves real-life problems involving length and angle relationships within the circle or the right-angled triangle.
3. finds unknown elements in general triangles using scale drawing or calculation methods, provided that the student is given the option of the method that he/she can use.
4. constructs a rigorous proof (not necessarily in two-column format) by stating properties, theorems, or corollaries involved in the solution.

Note: From Council of Ministers of Education, Canada. (2001). *School Achievement Indicators Program Mathematics Assessment: Criteria and framework*. Toronto, ON: Author.

Table 2

Measurement and Geometry Items on the Content Subtest of the 2001 SAIP Mathematics Assessment

Item Order	Item Type	Achievement Level	Target Ability
023	Multiple Choice	1	Conceptual
024	Multiple Choice	1	Conceptual
026	Multiple Choice	1	Conceptual
033	Short Answer	1	Conceptual
035	Short Answer	1	Procedural
036	Short Answer	1	Procedural
042	Multiple Choice	2	Conceptual
047	Multiple Choice	2	Problem Solving
049	Multiple Choice	2	Conceptual
053	Multiple Choice	2	Problem Solving
064	Short Answer	2	Procedural
065	Short Answer	2	Problem Solving
069	Short Answer	3	Problem Solving
074	Short Answer	3	Conceptual
083	Multiple Choice	4	Conceptual
084	Multiple Choice	4	Problem Solving
086	Multiple Choice	4	Conceptual
088	Multiple Choice	4	Conceptual
094	Short Answer	4	Procedural
096	Short Answer	4	Procedural
100	Short Answer	4	Problem Solving
101	Multiple Choice	5	Conceptual
105	Multiple Choice	5	Procedural

Item Order	Item Type	Achievement Level	Target Ability
108	Multiple Choice	5	Conceptual
109	Multiple Choice	5	Problem Solving
110	Multiple Choice	5	Problem Solving
125	Short Answer	5	Problem Solving

Note. Item classifications were provided by the Council of Ministers of Education, Canada.

Table 3

Ontario Students Responding to the Content Subtest of the 2001 SAIP Mathematics Assessment

		English-Language Test		French-Language Test	
Age	Starting Point	Number	Percent	Number	Percent
13	1	516	65.1	330	67.8
	2	214	27.0	114	23.4
	3	63	7.9	43	8.8
16	1	249	36.8	254	46.5
	2	234	34.6	190	34.8
	3	194	28.7	102	18.7

Table 4

Items Exhibiting DIF According to IRT or Mantel-Haenszel Approaches

	English					French					IRT Results			MH Results	
Item Number	Percent Attempted	Percent Correct	*a*	*b*	*c*	Percent Attempted	Percent Correct	*a*	*b*	*c*	χ^2_{diff}	df	*p*	$\chi^2_{M\text{-}H}$	*p*
13 YEAR OLDS															
Starting Point 1 (English *N* = 516, French *N* = 330)															
23	99.42%	83.91%	0.64	-1.34	0.26	98.18%	86.36%	0.32	-3.02	0.26	5.8	3	.122	4.1*	.042
33	97.09%	89.15%	0.76	-1.98		94.24%	77.58%	0.44	-1.92		19.3***	2	.001	10.2***	.001
35	96.12%	71.12%	0.66	-0.90		94.55%	73.33%	0.37	-1.84		7.5*	2	.024	4.9*	.026
Starting Point 2 (English *N* = 214, French *N* = 114)															
No Items															
Starting Point 3 (English *N* = 63, French *N* = 43)															
105	55.56%	25.40%	1.29	1.22	0.17	27.91%	4.65%	0.27	7.68	0.12	8*	3	.046	9.6**	.002
108	57.14%	22.22%	1.26	1.21	0.14	20.93%	11.63%	1.19	1.61	0.13	-2.5	3	*a*	4.1*	.043
110	52.38%	20.63%	0.81	1.91	0.16	18.60%	6.98%	1.09	1.95	0.12	-0.5	3	*a*	5.0*	.025
16 YEAR OLDS															
Starting Point 1 (English *N* = 249, French *N* = 254)															
33	95.18%	85.14%	0.68	-1.85		91.34%	78.35%	0.83	-1.21		4.9	2	.086	5.5*	.019

	English					French					IRT Results			MH Results	
Item Number	Percent Attempted	Percent Correct	*a*	*b*	*c*	Percent Attempted	Percent Correct	*a*	*b*	*c*	χ^2_{diff}	df	*p*	$\chi^2_{M\text{-}H}$	*p*
Starting Point 2 (English *N* = 234, French *N* = 190)															
47	98.72%	82.05%	0.44	-1.68	0.26	98.95%	88.42%	0.73	-1.74	0.25	5.8	3	.122	4.1*	.042
49	98.72%	74.36%	0.53	-0.84	0.25	98.95%	84.21%	0.68	-1.44	0.26	8.4*	3	.038	8.7**	.003
53	98.29%	73.50%	0.84	-0.59	0.24	99.47%	80.53%	0.71	-1.17	0.24	4.4	3	.221	4.1*	.042
94	62.82%	8.55%	0.87	2.12		45.79%	1.58%	0.72	3.59		8.5*	2	.014	8.8**	.003
Starting Point 3 (English *N* = 194, French *N* = 102)															
83	97.94%	67.53%	0.71	-0.18	0.26	100.00%	74.51%	0.76	-0.84	0.27	5	3	.172	4.9*	.026
84	94.33%	58.76%	0.99	0.15	0.21	96.08%	65.69%	0.51	-0.39	0.27	6.7	3	.082	6.2*	.013
101	85.57%	61.86%	0.71	-0.04	0.21	66.67%	38.24%	0.86	0.9	0.21	10.5*	3	.015	9.1**	.003

Note. *** $p < .001$; ** $p < .01$; * $p < .05$. *a* indicates that χ^2_{diff} is negative, so it is impossible to compute a probability.

Table 5

Comparison of IRT and M-H results.

Comparing IRT and M-H Procedures of DIF Detection										
					IRT (χ^2_{diff})			M-H (χ^2_{M-H})		
	Item Number	Item Type	Ability	Language Favored	.05	.01	.001	.05	.01	.001
13: Starting Point 1	23	MC	C	F				X		
	33	SA	C	E			X			X
							(19.3)			(10.2)
	35	SA	P	F	X			X		
					(7.5)			(4.9)		
13: Starting Point 3	105	MC	P	E	X				X	
					(8.0)				(9.6)	
	108	MC	C	E				X		
	110	MC	PS	E				X		
16: Starting Point 1	33	SA	C	E				X		
16: Starting Point 2	47	MC	PS	F				X		
	49	MC	C	F	X				X	
					(8.4)				(8.7)	
	53	MC	PS	F				X		

	94	SA	P	E	X				X	
					(8.5)				(8.8)	
16: Starting Point 3	83	MC	C	F				X		
	84	MC	PS	F				X		
	101	MC	C	E	X				X	
					(10.5)				(9.1)	
Total # of items				7 French	5	0	1	9	4	1
				7 English						

Note. MC = multiple choice; SA = short answer; C = conceptual; P = procedural; PS = problem solving, (.) = item *p* value

Figure Caption

Figure 1. SAIP administration with placement test and starting points.

Figure 2. English and French versions of Item 53 (released in *Report on Mathematics Assessment III*, CMEC, 2001).

Figure 3. English and French versions of Item 108 (released in *Report on Mathematics Assessment III*, CMEC, 2001).

Figure 4. English and French versions of Item 110 (released in *Report on Mathematics Assessment III*, CMEC, 2001).

Figure 5. Numbers of items attempted, 13- and 16-year-old students taking the English- and French-language versions and beginning at Starting Point 1.

Figure 6. Numbers of items attempted, 13- and 16-year-old students taking the English- and French-language versions and beginning at Starting Point 2.

Figure 7. Numbers of items attempted, 13- and 16-year-old students taking the English- and French-language versions and beginning at Starting Point 3.

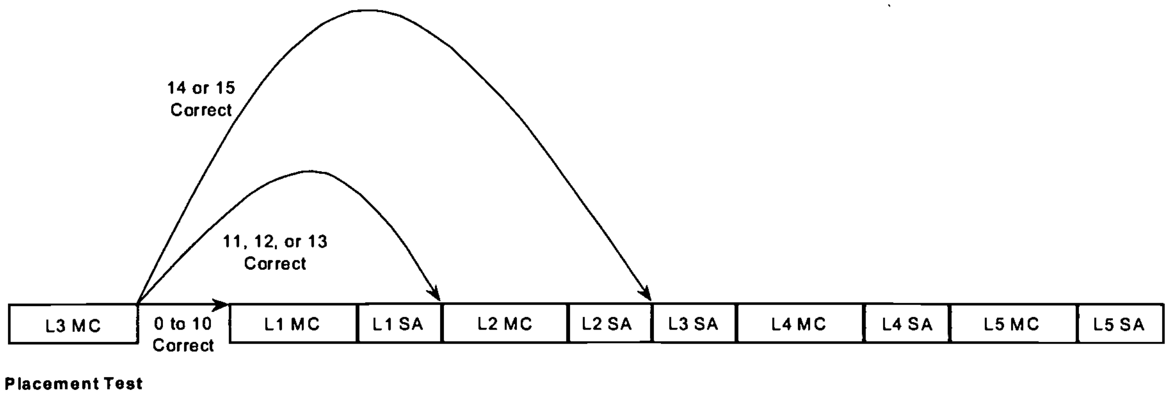
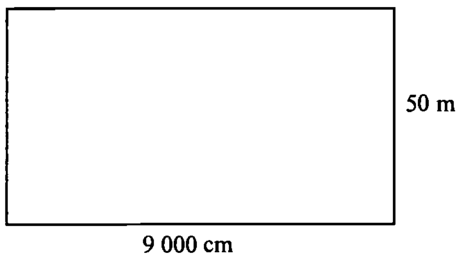


Figure 1. SAIP administration with placement test and starting points.

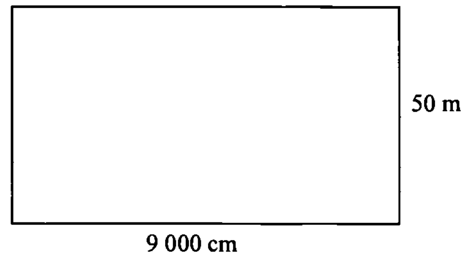
James wants to run the perimeter of a playing field, the dimensions of which are marked in centimetres and metres.



What distance will James cover by running once around this field?

A) 280 cm
B) 280 m
C) 18 100 cm
D) 18 100 m

Jacques court en suivant la ligne délimitant le contour d'un terrain de jeu dont les dimensions sont indiquées en mètres et en centimètres.

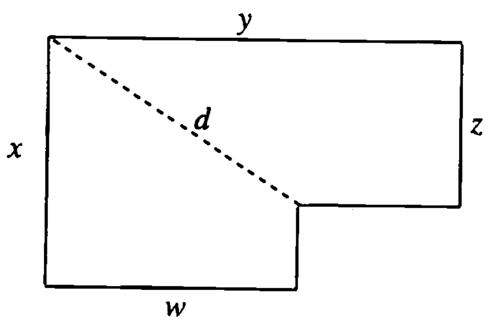


Quelle distance Jacques devra-t-il parcourir pour faire un tour complet du terrain de jeu?

A) 280 cm
B) 280 m
C) 18 100 cm
D) 18 100 m

Figure 2. English and French versions of Item 53 (released in *Report on Mathematics Assessment III*, CMEC, 2001).

The following diagram is a house plan. All corners are square.



What is the length of the diagonal *d* in terms of the variables given in the diagram?

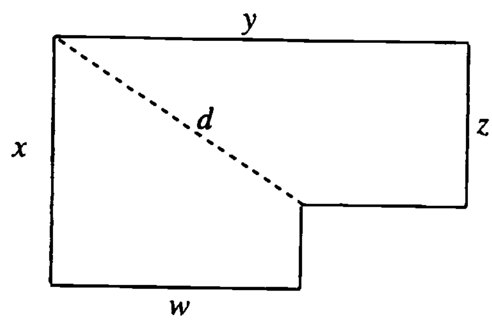
A) $d = \sqrt{x^2 + y^2}$

B) $d = \sqrt{w^2 + z^2}$

C) $d = \sqrt{x^2 + w^2}$

D) $d = \sqrt{y^2 + z^2}$

Le dessin ci-dessous représente le plan d'une maison. Tous les coins sont à angle droit.



Laquelle des expressions suivantes permet de calculer la longueur de la diagonale *d* en fonction des variables indiquées sur le dessin?

A) $d = \sqrt{x^2 + y^2}$

B) $d = \sqrt{w^2 + z^2}$

C) $d = \sqrt{x^2 + w^2}$

D) $d = \sqrt{y^2 + z^2}$

Figure 3. English and French versions of Item 108 (released in *Report on Mathematics Assessment III*, CMEC, 2001).

A drafting student must construct a symbol. The symbol consists of a circle of radius 30 cm and an inscribed equilateral triangle. A metallic wire is used to outline the perimeter of the triangle.
To the nearest centimetre, what is the length of metallic wire needed?
A) 90 cm
B) 156 cm
C) 180 cm
D) 188 cm

Un étudiant en graphisme doit reproduire un symbole. Ce symbole est formé d'un cercle de 30 cm de rayon et d'un triangle équilatéral inscrit. Un fil métallique délimitera le contour de ce triangle.
Quelle est, au centimètre près, la longueur du fil métallique?
A) 90 cm
B) 156 cm
C) 180 cm
D) 188 cm

Figure 4. English and French versions of Item 110 (released in *Report on Mathematics Assessment III*, CMEC, 2001).

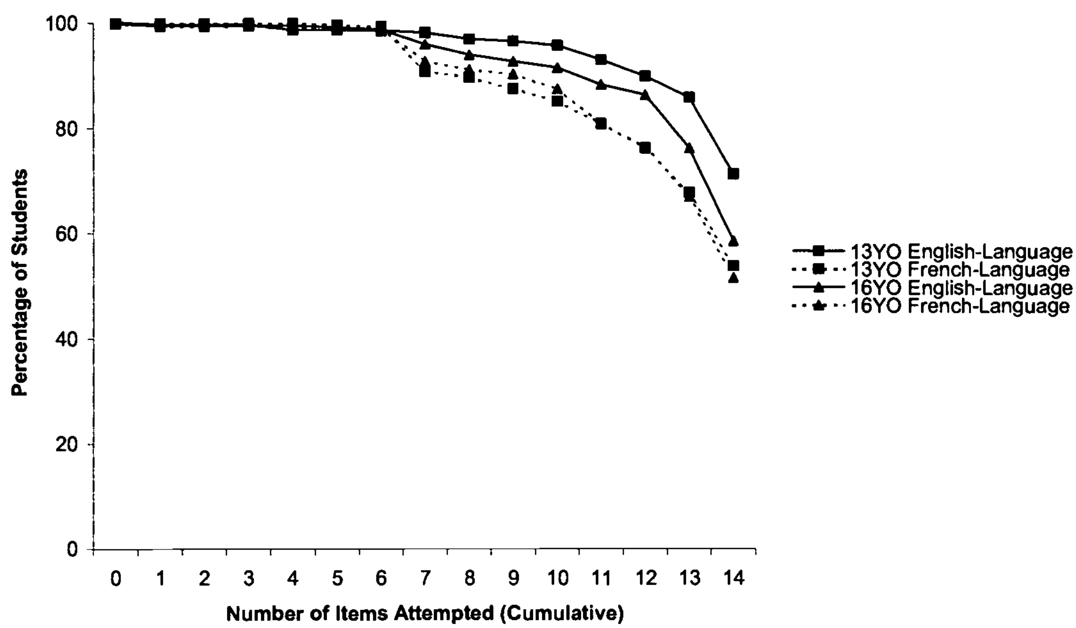


Figure 5. Numbers of items attempted, 13- and 16-year-old students taking the English- and French-language versions and beginning at Starting Point 1.

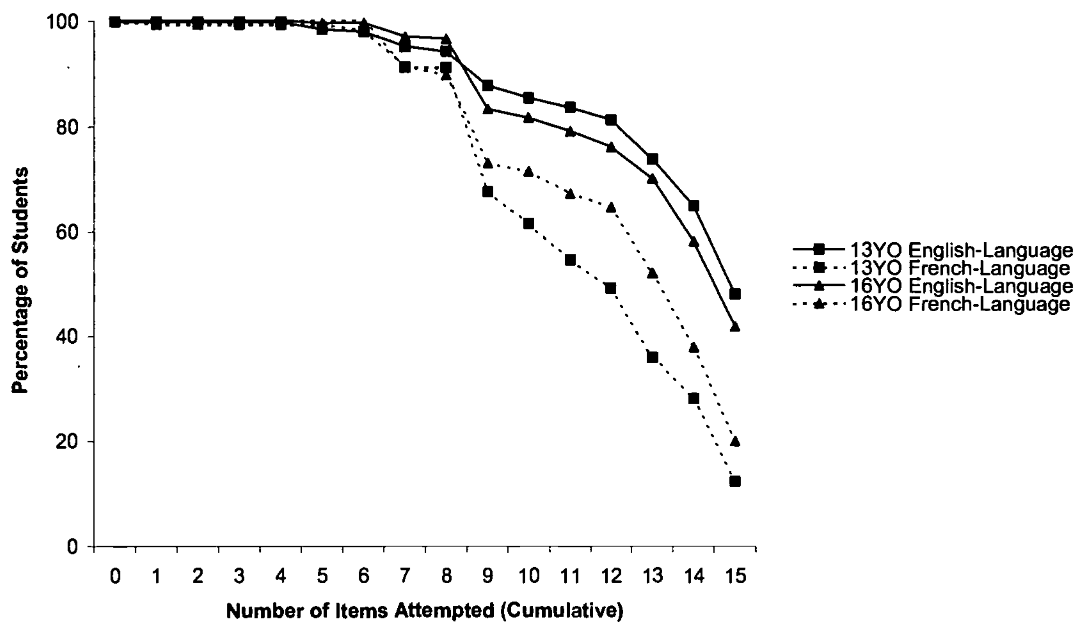


Figure 6. Numbers of items attempted, 13- and 16-year-old students taking the English- and French-language versions and beginning at Starting Point 2.

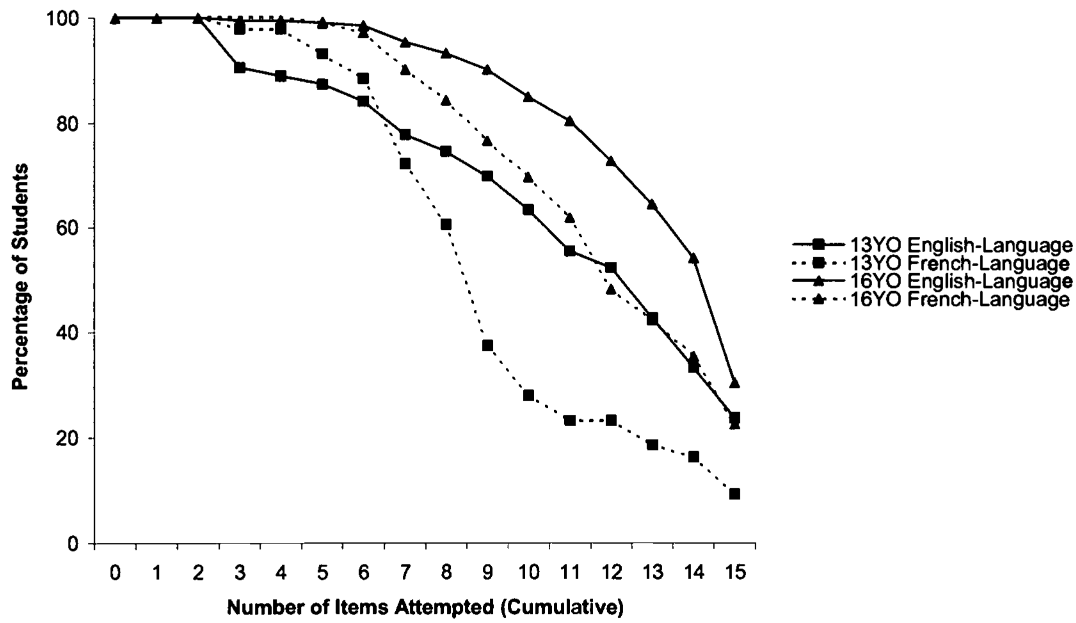


Figure 7. Numbers of items attempted, 13- and 16-year-old students taking the English- and French-language versions and beginning at Starting Point 3.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM034926

I. DOCUMENT IDENTIFICATION:

Title: CURRICULUM AND TRANSLATION DIFFERENTIAL ITEM FUNCTIONING: A COMPARISON OF TWO DIF DETECTION TECHNIQUES	
Author(s): BARNABAS C. EMENOGU AND RUTH A. CHILDS	
Corporate Source: CURRICULUM, TEACHING AND LEARNING DEPARTMENT ONTARIO INSTITUTE FOR STUDIES IN EDUCATION OF THE UNIVERSITY OF TORONTO	Publication Date: APRIL 23, 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]	Printed Name/Position/Title: BARNABAS EMENOGU (MR)	
Organization/Address: DEPARTMENT OF CURRICULUM, TEACHING AND LEARNING, OISE/UNIVERSITY OF TORONTO 252 BLOOR ST. WEST, TORONTO ON M5S 1V6	Telephone: 416-923-6641 ext 2743	FAX: 416-926-4744 [illegible]
	E-Mail Address: ~~bemenogu~~ bemenogu@oise.utoronto.ca	Date: May 11, 2003

(Over)