This study is an extension of a current ongoing project that examines data from the Pan-Canadian Assessment Program (PCAP) for Differential Item Functioning (DIF) in the English and French versions. In Canada, a bilingual country, tests are often developed in English first, then translated/adapted to French. The nature of such Test Adaptation and Translation process may cause the test to elicit social-linguistic differences between the language groups. Using the DIF items identified from the ongoing project, this research focuses on the English test-takers and their performances. Although some items in the test may favor the English takers over the French takers, difference in performance within the English group is also possible. As Canada is a multicultural country, the different levels of English spoken may cause variability in test performances, even among the items that suppose to benefit them as a group. To operationalize the idea of “different levels of English”, the demographic data provided by test-takers will be used to form groups and compare with each other using their performance on the DIF items identified. Demographic data reflecting their English fluency such as Canadian born or not, years lived in Canada, and language spoken at home will be used for comparisons. MANOVA and Factorial ANOVA will be performed to compare results of test takers from different groups. Interaction between factors will also be examined to determine their difference in performance.

Research Questions

This study aims to answer two major questions:
(1) Is there a within language group difference on test performance for items that were DIF to benefit that language group?
(2) If so, which factor(s) contribute to cause such a difference in performance?

If you have any questions, feel free to contact: Karen.Fung@ualberta.ca

Examining Language Proficiency, Test Performance, and Test Fairness using Data from the Pan-Canadian Assessment Program

Method

Our ongoing main project examines the existence of DIF between the English and French versions in the 2007 PCAP data. This proposed study serves as an extension by placing a microscope to focus on the DIF items and the group that ‘benefit’ from those items.

In this study, the English test-takers (Reading 1: n~5800; Reading 2: n~6000; Mathematics & Science: n~4600) will be categorized based on their responses on the questionnaire at the end of the test booklets. Questionnaire items providing clues on the test-takers’ English levels including (1) Canadian-born / not Canadian-born, (2) years lived in Canada, (3) language spoken at home, and (4) language most school subjects were taught will be used. These items will be the independent variables (IVs) used for the comparison to determine if performance on the DIF items (the dependent variables; DVs) from each of the Readings, Mathematics, and Science domain differ significantly between groups. Factorial Analysis of Variance (ANOVA) and Multivariate analysis of variance (MANOVA) will be used for comparing if there is any significant difference between groups. MANOVA is used because of the multiple subject measures for each student in the Mathematics and Science Booklet, and Factorial ANOVA is used for the Reading 1 and the Reading 2 booklets as each student was only measured in the Reading domain. If a significant difference is found, interactions between the group membership and the test subjects will also be examined. And finally, the main effects to which of the IV levels contribute to a significant difference in performance between groups can be determined.

Significance

This research is important to both the fields of language testing and psychometrics. It helps focus our attention on the issue of sociocultural difference that exist even within the same country, as well as highlighting issues related to test fairness and test validity. It also allows the test stakeholders and psychometricians to identify the variables that could affect the benefited group in any testing situation. When a set of items is categorized as DIF, not everyone in that benefited group may actually be gaining much. The variety of English levels found shows the importance of how items should be properly developed to ensure the questions were understood the way they were suppose to be.

Directions for Future Research

As this data set has a multilevel nature, Hierarchy Linear Modeling (HLM) could be used in future research to form a predictive relationship between variables at different levels and the students’ test performance. Future research may also examine whether demographic factors may contribute to the items in the PCAP to be DIF. For example, whether living in different provinces or even different communities in Canada would cause the items to be DIF due to the possible province-to-province, communities-to-communities difference in terms of cultures. Researchers may also extend to other countries in the world to determine if results found in this study can be generalized to other counties that are multilingual. For example, one may want to determine whether the same factors such as language spoken at home would play the same effect in affecting test performance for a certain language group. Additional factors not included in this study should also be considered in future studies as well. Examples of these multilingual countries may include Switzerland where four national languages are spoken, and Finland where two national languages are also spoken.

Funding for the main ongoing project and this current research were provided to Dr. EIAtia by: The Social Sciences and Humanities Research Council of Canada (SSHRC)
Travel Funding for this conference was provided by: The Myer Horwitz Graduate Student Travel Award