Early Child Development Mapping (ECMap) Project,
Community-University Partnership (CUP)

Applications of Differential Item Functioning (DIF) and Natural Language Processing (NLP) on Alberta's Early Childhood Development Instrument (EDI) Data

Hollis Lai, Ph.D.

&

Vijaya Krishnan, Ph.D.

2013 April 15

Faculty of Extension, University of Alberta, Edmonton, Alberta

Community-University Partnership
for the Study of Children, Youth, and Families

ECMap
Early Child Development Mapping Project Alberta

Contents

Summary

The report gives details of the validity of the Early Development Instrument (EDI), administered on kindergarten children in Alberta. The sample came from three waves of survey data over three years, 2009, 2010, & 2011 and children aged 4 to 7 years. This document gives details of the two methods, Differential Item Functioning (DIF) and Natural Language Processing (NLP) conducted to demonstrate possible teacher bias (by gender, English/French as second Language (ESL), and cultural/regional background of children) and to classify children with or without developmental challenges based on teachers' comments. It highlights some of the issues and challenges in adapting the tool in its present form, in a multi-cultural context.

The authors

Hollis Lai is a Curriculum Assessment and Evaluation Specialist with the Undergraduate Medical Education at the Faculty of Medicine and Dentistry and Vijaya Krishnan is a Research Scientist with the Early Child Development Mapping (ECMap) Project at the Community-University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta.

Acknowledgments

The authors are responsible for the choice of the methods, the presentation of the results and for the opinions expressed therein, which are not necessarily those of the University they work for or Alberta Education who is responsible for collecting the EDI survey data. The authors acknowledge the contributions of Dr. Susan Lynch, Director of the ECMap Project in providing her support in carrying out this project smoothly. The support of the entire ECMap team and Mr. Gregory Chao, in particular is also appreciated. Partial funding for this research was provided to Dr. Krishnan by the Faculty of Extension, University of Alberta, Edmonton, Alberta, Canada (RES0014743).

Copyright

Citation

Lai, H. & Krishnan, V. (2013). Applications of Differential Item Functioning (DIF) and Natural Language Processing (NLP) on Alberta's Early Child Development Instrument (EDI) data, Early Childhood Mapping (ECMap) Project, Community-University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta, Canada.

1. Introduction

Canada is a multicultural society whose ethnic and cultural composition has been shaped over time by people of different nationalities, as well as by the original inhabitants of the country, the Aboriginal peoples. The sources of immigrants to Canada have changed in recent decades with a majority of new comers to the country arriving from non-European countries, making it one of the most ethnically diverse nations in the world. The ethno-cultural diversity compels us to rethink the developmental needs of young children of immigrant families so that the disruptive effects of challenges and stresses, if any, immigrants face in their everyday life can be minimized. It is timely to assess measures such as the Early Development Instrument (EDI) so that they can be culturally appropriate.

Questions on cross-national application of psychological tests are common in cross-cultural research. Concrete examples of the problem are found in the work of Butcher & Garcia (1978), and specifically when they asked the questions: how do we know that a tool developed in a general population (dominant culture) will be applicable to populations in a different ethnic or cultural context? In norm-referenced test situations, how can we assume that a test score on the original population means the same thing in another cultural context? That is, if the instrument involves trait- or skill-based constructs (as in the EDI), there is no justification for its continued use without post-hoc analyses and empirical validation across different ethnic and cultural communities. The use of an unreliable tool can lead to an unfair assessment of outcomes, due to the tool itself and the errors that occur at various levels of its execution. It goes without saying it is important to study human behavior in the context in which it occurs or take an emic approach in order to reduce the bias in interpreting results.

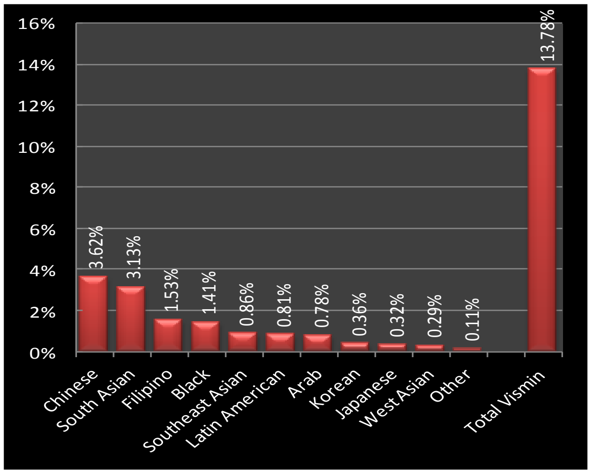
Modern psychometric methods enable cross-cultural analyses of biases of measures, giving test developers some scientific evidence for a successful test adaptation. The purpose of this study is to apply two such methods for analyzing responses to the EDI survey in the context of Alberta. The EDI being a teacher response survey administered to capture a comprehensive picture of assessment of young children's development at the community level, we aim to address its effectiveness so that it is suitable for adaptation in diverse communities. The analysis involves two separate methods, one describing the use of Differential Item Functioning (DIF) to study response biases in items for specific cultural communities, and the other demonstrating the use of Natural Language Processing (NLP) to analyze differences in teachers' comments.

2. Why consider cultural diversity in early childhood development research?

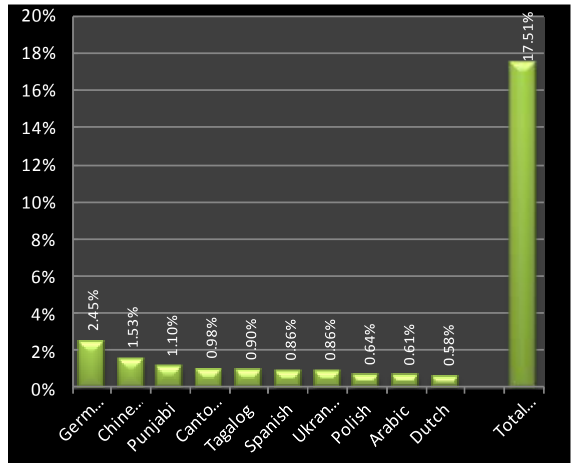
Worldwide, Early Childhood Care and Education (ECCE) is a priority of national governments in recognition that investing in children is a critical step in broader development. Accordingly, almost all societies make provisions for children's basic needs and initial learning from the very early age to ensure their survival and also promote physical, cognitive, social, and emotional developmental outcomes. A lot of progress has been made in recent years in improving early learning opportunities, but still far to go in breaking the cycle of developmental inequities that has dominated the lives of millions of children and families in adverse circumstances.

A remarkable feature of ECCE is that it is the product of European and North American culture, which represents only a minority of children in some multicultural societies. As a result, the benefits from ECCE bypass government and public efforts that support families and young children from visible minorities. In the report, *Strong Foundations: Early Childhood Care and Education,* UNESCO (2007) pointed out the often overlooked advantages of mother tongue-based multilingual education in the early years, based on empirical research across cultures (Ball, 2010). First, when children are offered opportunities to learn in their mother tongue, they are more likely to succeed in school (Kosonen, 2005). Second, parents are more likely to communicate with teachers and participate in children's education (Benson, 2002). Finally, and more importantly, mother tongue-based education is more likely to benefit children from rural, ethnic, and indigenous communities (Hovens, 2002). Most preschool programs neither offer support to develop competence in their mother tongue for minority and indigenous children nor value the ethnic or cultural backgrounds of children and their mother tongue (Ball, 2010).

Alberta is home to a growing number of immigrant families with young children. The communities are increasingly becoming more ethnically diverse than in the past with Chinese and South Asians representing a majority of the visible minority population (Figure 1). Visible minorities accounted for 13.8 percent (up from 11 percent in 2001) and allophones accounted for 17.5 percent of Alberta's population in 2006. The Aboriginal population has risen as well with six percent reporting to be of Aboriginal ancestry and one in three persons being 15 years of age and under (Statistics Canada, 2008). These statistics compel a paradigm shift to introduce culturally sensitive ECCE design and delivery.[1]



Visible minority



Allophones

Figure 1: Percent visible minority and allophones (people whose mother tongue is neither English nor French), Alberta, 2006

Source: Statistics Canada, Census 2006

The proximal causes of poor child developmental outcomes may vary across time and space. However, in view of the high cost of poor child development, both economically and in terms of potential for developmental risks, there is a need for testing whether or not the items in the EDI are psychometrically good. It may be that the instrument was developed primarily with a focus on behavioral indicators of early child development that was based on Anglo-American realities and values, denying the right to identity the knowledge systems of a growing number of Albertans and, consequently, threaten the predictive validity of outcomes. An in-depth analysis of the instrument is especially timely given the slow progress in meeting targets articulated in the UNESCO's (1990) *Education for All* goals, primarily of ECCE. This report explores the ability of items in the EDI to discriminate between major language groups, along with other child characteristics, such as sex.

[1] Annually, the province of Alberta invests close to $575 million in public funding for ECEC, with the Ministries of Children and Youth Services and Education providing the bulk of these monies (Government of Alberta, 2010; see also, The Muttart Foundation, 2010)

3. Data Analysis and Methods

The discussion here is based on *The Early Child Development Instrument (EDI): An item analysis using Classical Test Theory (CTT) on Alberta's data*, which is perhaps the only document of its kind within the context of Alberta (Krishnan, 2013).

3.1. The Early Development Instrument (EDI)

The EDI is a tool to assess kindergarteners' development in the five areas of development: physical health & well-being, social competence, emotional maturity, language & thinking skills, and communication & general knowledge. The tool is designed to be universal enough to be relevant to most preschoolers around the world, allowing an assessment and an overview of the five key areas with no component of screening, yet constructed from the perspective of a Eurocentric epistemology. The multidimensional EDI is geared to provide a methodology and a framework for communities to address developmental difficulties in children at a macro-level. Specifically, the EDI is a survey-based thematic tool primarily designed to assist and target communities at a local level, although data are collected at an individual level.[2]

A brief description as to how the teacher responses to survey questions are turned into numerical values or how the component scores are built is in order. Figure 2 visually illustrates the steps to aggregation into component scores. Once the data are collected, they are checked following a rigorous quality control process. In order to arrive at one pool of data from multiple waves, they are merged keeping the original codes intact. The three sections of the questionnaire, A, B, & C provide the information to create component scores. Items from each of the three sections (in different combinations) are then themselves combined in order to yield the component values, assigning equal weights to individual items.

[2] The tool is individually administered and norm-referenced with a strong emphasis by its developers on the importance to document the results to larger contexts by aggregation and analyses at the neighbourhood, sub-community and/or group level (based on age, sex, or ethnic characteristics).

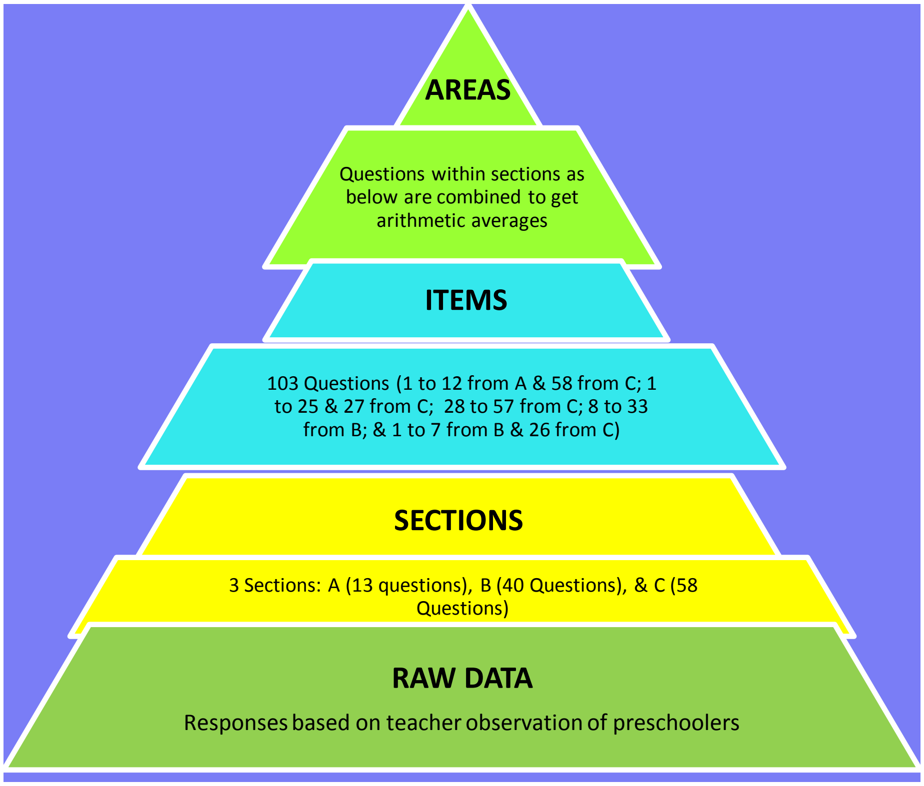


Figure 2: How EDI survey data are converted into five developmental area scores?

In summary, the EDI consists of five developmental areas with 103 questions to be answered on all five with 13 on physical health & well-being, 26 on social competence, 30 on emotional maturity, 26 on language & thinking skills, and eight on communication & general knowledge. The 103 questions associated with the five areas are referred to as items, in all our discussions. Thus, the composite of physical health & well-being is associated with 13 items, social competence with 26, and so on. The five components are presented in Figure 3 below. In this report, we use the terms component and domain interchangeably. The order is not intended as a ranking of the components, but they are usually presented in this sequence.

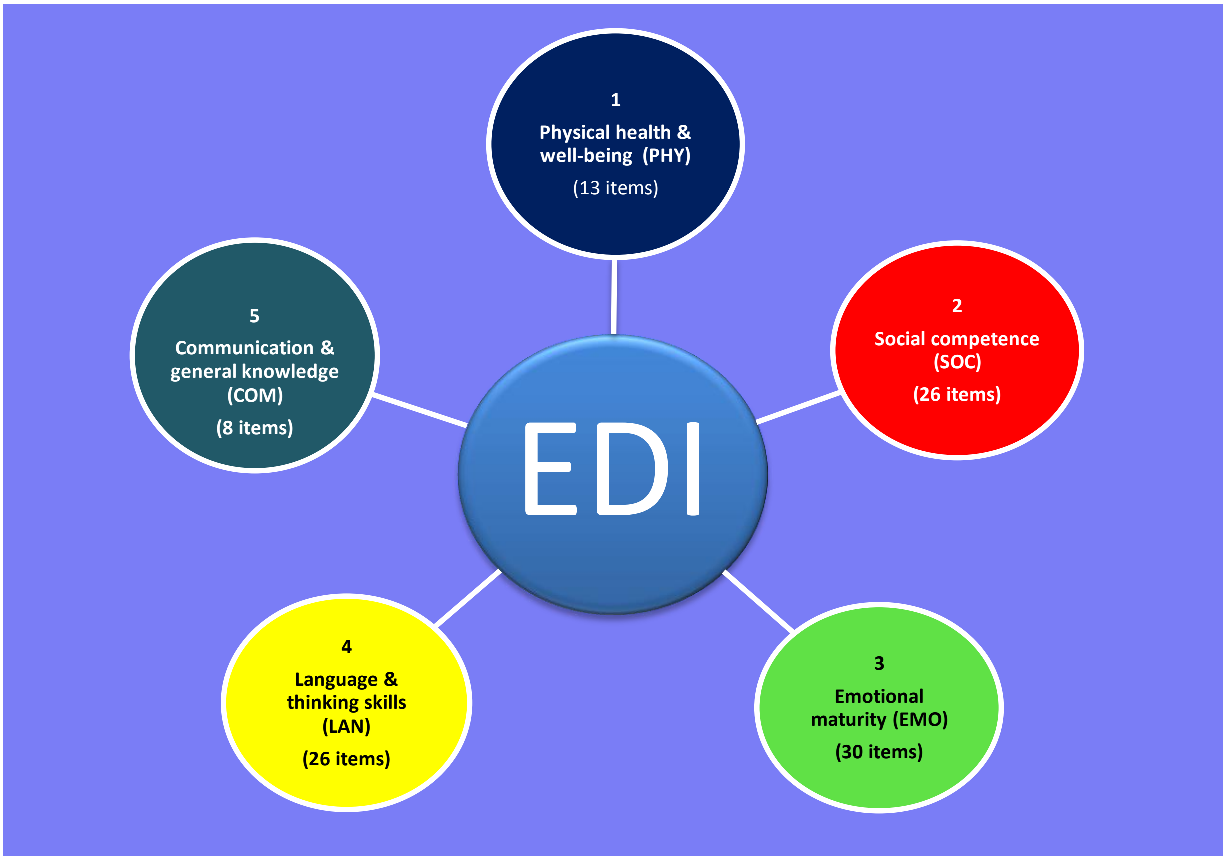


Figure 3: The EDI components/domains

3.2. Sample

Since its development in 1999 by the Offord Centre for Child Studies at McMaster University, the EDI is being utilized in a growing number of countries and all provinces and territories within Canada. The discussion that follows is from the analysis of the EDI survey questions, administered by kindergarten teachers across Alberta, through a collaborative effort led by Alberta Education and the Offord Centre. The Early Child Development Mapping Project (ECMap) (formulated in 2009) affiliated with the Community-University Partnership (CUP) at the University of Alberta is responsible for mining the data and developing an inter-community snapshot of developmental patterns of preschoolers.

The data for this study cover three waves (2009, 2010, & 2011) of data collection on preschoolers in Alberta, and represent 52,498 kindergarteners, ranging from age 4 to 8 years, in general. From this population, a sample of 40,484 valid EDI responses from children age four to seven was analyzed, using two major analytic procedures, as described below.

3.3. Analytical procedures

Given the variation in population composition and its effects on developmental outcomes at a macro-level, the present study employed two procedures in analyzing the data with the two important purposes:

- to investigate cultural response biases of teachers; and
- to enhance existing analytic processes to fully utilize the qualitative information (teachers' comments) as captured in surveys.

Since the two purposes require different segments of data, the study was broken down into two component parts, Part 1 describing the use of DIF to study response biases in items for specific cultural groups, and Part 2 demonstrating the use of latent semantic analysis, NLP to analyze differences in teachers' comments.

3.3.1. Differential Item Functioning (DIF)

As earlier noted, EDI is a teacher completed survey designed to provide population-based assessment of young children's learning across multiple domains for guiding policy development and intervention strategies (Janus, 2006). Unfortunately, the validity and reliability of the five domains in EDI are given very little attention (e.g., Hymel, LeMare & McKee, 2011; Janus, Brinkman & Duku, 2011), even though there can be possible sources of bias. As the instrument is administered to each child, results of the EDI are reported at the community level, where gender composition, cultural differences, and language often differ across groups and communities. That is, developmental outcomes may vary so radically between two cultures, such as Spanish and Turkish that, if the two were treated as one, they are likely to lose construct validity (Cook & Campbell, 1979). To measure children's development independent of demographic variations, items on the EDI need to be free from such biases. However, to our knowledge, only one study has investigated this issue (Guhn, Gadermann, & Zumbo, 2007), in a systematic way. In Part 1, we investigate whether cultural response biases exist for the EDI, within the framework of DIF.

DIF analyses can yield information about response biases, which is the first characterization of fairness, according to the psychological organizations, such as APA, AERA, and NCME (1999). According to Sireci & Allalouf (2003), analyses of DIF attempts to sort out whether item impact – a significant group difference on an item – is due to overall group differences in proficiency or due to item bias. The development and application of DIF detection methods is prominent in test development as it reflects, in large part, a response to the legal and ethical needs to assess examinees without error or bias. Analyzing for biases in survey responses is a relatively novel application. In general, to conduct DIF analyses, examinees are first divided into two groups, a *reference* and *focal group*. Second, a comparison of item responses is conducted against an overall performance on the instrument. Third, the comparison is compounded for respondents of a specific group. Finally, using statistical procedures, items are examined to see whether or not a given item show different pattern for the two groups. An item exhibits DIF when respondents from the reference and focal groups differ in terms of the probability of positive responses to a given item, after controlling for the measure of ability derived from the overall instrument.

DIF analysis on EDI involves multiple layers of complexity. First, the instrument contains five components (domains), and the scores are reported for each component separately. Second, each component contains a different composition of items and also different item length. For example, some components have only dichotomous items, while others have a combination of scaled items as well as dichotomous ones. The response scale used for each component dictates the DIF analytic methods as well as the overall score of responses. Third, depending on the group criterion, say English as a second language, the requirement of a reasonably good sample size has been difficult to meet. With these three layers of complexity in utilizing DIF, the analysis of EDI results require a sophisticated design that require multiple interpretations for every item.

To determine whether item responses on the EDI contain cultural biases, for the DIF part of the analysis, three pairs of comparisons were made based on three child characteristics: gender (female vs. male), English as a Second Language (ESL) vs. English as a first language, and a specific cultural or regional background (South Asian language speakers vs. Non-South Asian language speakers). The three comparisons were selected to demonstrate: (a) gender DIF – broader comparison of one sex over other; (b) ESL DIF – broader comparison of native language speakers to those from a different linguistic background, and (c) culture DIF – broader comparison of Non-South Asian children to South-Asian children. Please note that South Asian children were identified via teachers' responses of primary languages associated with children of South Asian origin.

Ordinal logistic regression was used to detect DIF (Swaminathan & Rogers, 1990). An estimated R^2 using likelihood ratio differences was calculated to classify DIF items that contained ordinal and dichotomous responses. The analysis was conducted in R (R Core Team, 2012). DIF using ordinal logistic regression follows the logistic function,

$\theta_n = \frac{1}{1+e^{(\beta 1 X1 + \beta 0)}}$ Where, θ stands for all the possible responses for a given item; $\theta \epsilon \{1,2 \ldots n\}$.

For example, if items in the EDI have the options, 0, 5 or 10, then, an individual β_0 is assigned to each outcome, and responses are assumed to be scalable onto a common scale of probability. DIF is represented by the difference between variance accounted for by a model estimated using the total scale score $(\beta 1 X1 + \beta 0)$ compared to a logistic model estimated using total scale score and group criterion $(\beta 1 X1 + \beta 2 X2 + \beta 0)$. Using an estimated variance accounted for by the model from Chi-Square likelihood of fit for data with the estimated model, the resulting difference (ΔR^2) between the two models represents the influence of a group category on the item response. This influence can also be interpreted as the response bias for a given criterion, where such information can be used to interpret the magnitude of the bias in response. Using a three-class DIF criteria proposed by Jodoin & Gierl (1999), we categorized response bias categories into three: minimal $(\Delta R^2 < .035)$; moderate $(\Delta R^2 < .07)$; and serious $(\Delta R^2 \geq .07)$. In addition, a comparison of non-uniform DIF resulting effect from the interaction of group criterion and performance is also presented. In all, each item on the EDI produces two ΔR^2 for each of the three groups to be compared. This analysis is carried out for all five domains of the EDI.

Descriptive Statistics

A statistical summary of each of the five domain of the EDI is first presented for the three waves of data as in Table 1. A comparison of results between years did not yield a significant difference.

Table 1: The five domains and their mean scores

Year	N	PHY	SOC	EMO	LAN	COM
2009	9641	8.46	8.20	7.86	8.30	7.26
2010	21976	8.62	8.27	8.02	8.33	7.47
2011	20881	8.54	8.28	7.96	8.25	7.51

In Table 2 are presented the conditional summary of each component for each of the three DIF grouping criteria. The three DIF categories posed an increasing level of specificity when comparing to a minority population. With increasing specificity, the sample size for the focal group decreases. However, from the large sample of respondents, the minimum sample size requirements were met for all DIF comparisons.

Table 2: Percentage distribution of children by sex, language, and region

Category	Label	N	Percentage
Gender	Female	25199	48%
	Male	27262	52%
ESL	Yes	6723	13%
	No	43259	87%
South Asian	Yes	1617	3%
	No	48365	97%

Detection of DIF

Recall that DIF is independent of children's overall performance on the scale. Because the responses are based on teachers' observation of children's behaviour, items that exhibit DIF likely represent response biases of teachers for a given population. In Table 3 are given the level of DIF for items in communication & general knowledge component, while remaining DIF results for all other categories and populations are reported in Appendix A.

Table 3: DIF results for items in the communication & general knowledge area of EDI

DIF Category	Item	DIF		R		
		Uniform	Non-Uniform	Sub Score Only	Sub Score DIF CAT	Sub Score DIF CAT Interaction
Gender	1: effective use-English	0.00	0.02	0.37	0.37	0.38
Gender	2: listens-English	0.00	0.02	0.37	0.37	0.39
Gender	3: tells a story	0.00	0.02	0.39	0.39	0.41
Gender	4: imaginative play	0.00	0.02	0.37	0.37	0.40
Gender	5: communicate needs	0.00	0.02	0.37	0.37	0.38
Gender	6: understands	0.00	0.02	0.36	0.36	0.38
Gender	7: articulates clearly	0.00	0.02	0.31	0.31	0.33
Gender	8:knowledge about world	0.00	0.01	0.27	0.27	0.27
ESL	1: effective use-English	**0.07**	**0.12**	0.28	0.35	0.40
ESL	2: listens-English	0.03	**0.07**	0.30	0.34	0.37
ESL	3: tells a story	**0.04**	**0.10**	0.31	0.35	0.41
ESL	4: imaginative play	0.01	**0.06**	0.32	0.33	0.38
ESL	5: communicate needs	0.03	**0.09**	0.29	0.32	0.39
ESL	6: understands	0.03	**0.08**	0.29	0.32	0.38
ESL	7: articulates clearly	0.00	**0.05**	0.26	0.27	0.31
ESL	8:knowledge about world	0.01	**0.04**	0.20	0.21	0.24
South Asian	1: effective use-English	**0.04**	**0.11**	0.25	0.29	0.36
South Asian	2: listens-English	0.02	**0.04**	0.42	0.44	0.47
South Asian	3: tells a story	**0.05**	**0.15**	0.38	0.42	0.53
South Asian	4: imaginative play	0.00	**0.03**	0.37	0.38	0.40
South Asian	5: communicate needs	0.02	**0.10**	0.25	0.27	0.35
South Asian	6: understands	0.03	**0.09**	0.25	0.28	0.33
South Asian	7: articulates clearly	0.00	**0.10**	0.37	0.37	0.46
South Asian	8:knowledge about world	0.01	**0.03**	0.38	0.39	0.40

Conclusions drawn from DIF analysis

Using a three-class DIF criteria proposed by Jodoin and Gierl (1999), our study found no significant item biases for all three comparison groups in four of the five EDI domains (physical health and well-being, social competence, emotional maturity, and language and cognitive development). However, we found five of the eight items in the communication skills domain to exhibit severe non-uniform DIF for both ESL and South Asian children. While the majority of the EDI item responses are robust against language and cultural differences, teachers' assessment on children's communication skills is less reliable for children of different linguistic and cultural backgrounds, where items in that domain may be confounded with language learning issues. The results of our study suggest the need for more investigation to better guard against unintended item biases in providing information on communications skills and general knowledge.

Studies involving DIF on minority populations are instrumental in eliminating biases in test development. However, applications of DIF on survey items, specifically on early childhood development are less prominent. DIF, as demonstrated in this study, can be applied to determine response biases to alleviate the potential problems that might be encountered by researchers and policy makers when adapting instruments from another language that is different from the test-takers' dominant language. Results from this study can be used to refine instrument development, guide item selection for shortened versions of the instrument, and more importantly, refine evidence for informing early childhood development in communities and/or neighbourhoods.

There are few limitations to the results of this study. First, small negative values have been found for some of the items; the addition of a DIF category as a predictor variable actually decreased the proportion of variances accounted for in the logistic regression. This artifact is likely a result of the large discrepancy in sample sizes as the focal population becomes more specific, where the McFadden's R^2, is only an approximation of the proportion of variances accounted for. In other words, the negative values are likely small rounding errors compounded over the large sample set. Second, similar to DIF studies in educational measurement, results of the DIF study cannot conclusively determine the causal reasoning for such items exhibiting DIF, but rather provide evidence that such items on the EDI elicit biased responses from learners of a given population. Further, since the EDI is a teacher completed instrument and not a self-completed one, an additional assumption has to be made that all teachers in general elicit the same level of bias with regard to the items.

As a closing remark, the EDI collects comprehensive information about preschoolers' development within communities. But if teacher responses are biased toward demographic characteristics that are known to differ between communities, then the collected data may not be an accurate representation of the community. Our study suggests modifications are required to the communication skills domain in order to ensure the accuracy of the portrait of school readiness within communities. In the second part of our analysis, we present a demonstration of how qualitative comments from the EDI can be analyzed using modern NLP techniques.

3.3.2. Natural Language Processing (NLP) of teachers' comments

The EDI is administered across provinces and territories in Canada, demonstrating the societal desire to make success in school the norm for each and every child, regardless of their different life experiences and orientations. With the information collected at the community level, policy makers are able to address disparities related to developmental disadvantages of young children, and consequently guide resources for learning interventions. There is a general consensus in the literature that the EDI in general has construct validity (e.g., Janus et al., 2011). Specifically, using the common source of construct-related validation information, domain structures of EDI have been found to be representative of the data (Forer & Zumbo,2011), with domains tested using confirmatory factor analytic approaches (Hymel, LeMare & McKee ,2011; Janus et al., 2011). However, in some cases, it has been found that the loadings of items on factors differ across provinces (e.g., Krishnan, 2010; 2013), although there are those who report validity evidence across different programs (e.g., Santos, Brownell, Ekuma, Mayer & Soodeen, 2012). In such cases, however, the reporting of the EDI, including its evidence of validation and reliability, has largely focused on the quantitative aspects of the instrument. That is, one validation design that is often neglected in EDI research is that based on the qualitative data reported by teachers.

Qualitative feedback is collected on the EDI via one open prompt for teachers to provide any comment they deem appropriate to the child. Responses are written by teachers of the learner, which then is typed in the data entry process. With over 52,000 responses collected over a period of three years, such information is difficult to summarize as thousands of comments are available for children in any community, causing researchers to often overlook crucial information teachers are trying to convey. For example, Krishnan, Huaitng & Babenko (2011) provided a highlight of comments in their report, where a sample from it was then summarized under the five components in EDI. Clearly, however, no study has ever investigated the qualitative feedback collected in the EDI as we have attempted here.

Emerging advances in NLP have enabled new methods for analyzing the text material. Technology now allows for the identification of similarly worded item pairs (Lai & Becker, 2010), generation of text items from a corpus of text (Karamanis, Ha, & Mitkov, 2006), and the use of semantic similarity to estimate test item difficulty (Belov & Knezevich, 2008). These studies suggest the emergence of a methodology is currently being evolved to analyze qualitative and subjective data using state of the art technologies. The purpose of this part of the study, then, is a demonstration of how comments from the EDI can be analyzed in order to produce meaningful information using a new technique. More specifically, in this part of the study, we present an alternative method involving the use of NLP to provide more information from qualitative feedback of the learners.

The process of NLP involves three stages of development. First, comments are processed into particular formats. In this stage, each comment is broken down into a set of features, which are then compiled across all comments to determine a unique set of information. Second, the comments and features are used to predict the classification of children based on, say the vulnerability status. The accuracy of the classification process may not be paramount, but results from this analysis could provide a summary of what features of a comment is a predictor of the vulnerability status. Finally, features with the classification data are used to summarize vulnerable and non-vulnerable children. The analysis was conducted using lightSIDE (Mayfield & Rose, 2012), an NLP software and the LSA package from R, was used.

Data used for NLP

From a sample of 40,484 valid EDI responses, this part of the study only utilized responses with comments. A total of 3759 responses, or 9% of the total number of responses, were used for this analysis. Table 4 below describes the proportion of vulnerable children in this sample, determined based on a conventional cut-score on their performance. That is, children who score at or below the 10th percentile threshold of the normative group are considered experiencing great difficulty (vulnerable as has always been referred to in the literature). Thus, if the 10th percentile cut-off value is 7.0833 (as in the case of the physical health & well-being area), then all those children who score at or below 7.0833 are considered experiencing great difficulty or vulnerable in that domain.

Table 4: Proportion of vulnerable children based on cut-scores

	PHY	SOC	EMO	LAN	COM
Not Vulnerable (N)	2605	2858	2716	2871	2533
Vulnerable (N)	1128	877	997	857	1196
Percentage	30%	23%	27%	23%	32%

The percentages based on our classification of the sample with comments are significantly greater than the expected rate in the population (~10%), of course, as a result of the fact that the cut score is set using a distributional assumption. However, representativeness of the population should not adversely affect our analysis as the classification of vulnerability is used only to link the corresponding comments, hence does not require proportional representation.

NLP Results

In the first step of the analysis, comments from the EDI were broken down into a set of features. In this process, 7,631 unique features were extracted from the 3,759 comments. The four types of features identified in this process were:

- ***Unigram***, a unique word or type of word (e.g., noun, adjective) that appears in the comment;
- ***Bigram***, a unique consecutive word pair or word type pair that appears in the comment;
- ***Trigram***, a unique consecutive segment of three words that appears in the comment; and
- ***POS bigram***, a pair of words that may appear in any order in a segment of three words.

These features were counted and collected for all comments, where a majority of the overlaps occur with word types. Moreover, the occurrence of a feature was only counted once per comment. After comments were broken down into features, a classification process was used to determine which features determine the classification of vulnerability for each category. To undertake this task, a Bayesian network was used to compute the classifications and relationships of each comment. The entire sample was used for classification where the sample was divided into 10% segments to arrive at a convergence. A unique network was trained for each of the five categories of EDI. After training, each feature had a weight to calculate the likelihood of a comment being described as vulnerable or not. The results of this classification exercise are presented in Table 5 below.

Table 5: Classification results based on NLP

Value	PHY	SOC	EMO	LAN	COM
Correctly Classified (Percent)	2451 (65%)	2692 (72%)	2560 (68%)	2735 (73%)	2538 (68%)
Incorrectly Classified (Percent)	1308 (35%)	1067 (28%)	1199 (32%)	1024 (27%)	1221 (32%)
Kappa	0.19	0.25	0.22	0.24	0.26
RMSE	0.45	0.42	0.43	0.40	0.43

From the results extracted from the classification, the proposed method was able to match a high of 73% of the classification results from the cut score for the language & communication domain, and a low of 65% for the physical health & well-being domain. Although the classification results yielded relatively low agreement rates with the classification responses, results from the comments still could yield some features that can consistently classify vulnerable learners.

After features were used to classify vulnerability, results from this analysis provided a list of keywords that was highly discriminative on identifying vulnerable learners. Component- or domain-based results are provided below (Table 6 through 10). For each table of features, *predictor 0* describes non-vulnerable children, whereas *predictor1* describes the vulnerable ones. Kappa and Precision describe the classification of the specific feature, whereas the Hits describe the number of occurrence of a feature appearing in a comment. The number of Hits is further divided into the number of times it appeared on non-vulnerable children (Hit 0) and that appeared on vulnerable children (Hit 1).

Physical health & well-being (PHY)

To classify vulnerability based on physical health & well-being, comments of non-vulnerable children are often described as: outgoing, eager, able to count to 10, and descriptions that are indicative of their interaction with others (Table 6). In contrast, comments of vulnerable children often contain features that describe them as: puf-funding, removed from some place, and mentioning autism or autistic like symptoms.

Table 6: Vulnerability based on features of health & well-being

Feature	Predictor	Kappa	Precision	Hit	Hit 0	Hit 1
count_to_10	0	0.01	0.98	43	42	1
eager	0	0.01	0.93	43	40	3
very_bright	0	0.01	0.93	42	39	3
question_29_curriculum	0	0.01	0.97	34	33	1
section_b_question	0	0.01	0.97	31	30	1
doing_very	0	0.01	1.00	22	22	0
is_reading	0	0.00	1.00	19	19	0
outgoing	0	0.00	1.00	17	17	0
advanced	0	0.00	1.00	17	17	0
is_repeating_kindergarten	0	0.00	1.00	16	16	0
achieving	0	0.00	1.00	16	16	0
others_EOL	0	0.00	1.00	15	15	0
very_well_in	0	0.00	1.00	15	15	0
funded	1	0.01	1.00	9	0	9
diagnosed_with_autism	1	0.01	1.00	9	0	9
eyes	1	0.01	1.00	8	0	8
of_puf_funding	1	0.01	1.00	7	0	7
Reserve	1	0.01	1.00	7	0	7
Stem	1	0.01	1.00	6	0	6
was_removed_from	1	0.01	1.00	6	0	6
occupational_therapist_and	1	0.01	1.00	6	0	6
stem_from	1	0.01	1.00	6	0	6

Social competence (SOC)

In Table 7 is presented summaries of the features that are highly discriminative of social competence vulnerabilities. Features that are linked to children who are not vulnerable in the social competence area are often described with the terms: quickly, very ready, eager, or have attended junior kindergarten or Montessori school. Conversely, features that are associated with social competence vulnerabilities include: struggles, pediatrician, foster care, and syndrome.

Table 7: Vulnerability based on features of social competence

Feature	Predictor	Kappa	Precision	Hit	Hit 0	Hit 1
quickly	0	0.00	1.00	29	29	0
very_ready	0	0.00	1.00	28	28	0
her_mom	0	0.00	1.00	27	27	0
attended_junior_kindergarten	0	0.00	1.00	27	27	0
Montessori	0	0.00	1.00	24	24	0
academic_skills	0	0.00	1.00	22	22	0
willing_to	0	0.00	1.00	21	21	0
is_reading	0	0.00	1.00	19	19	0
eager_to_learn	0	0.00	1.00	19	19	0
has_come	0	0.00	1.00	19	19	0
Syndrome	1	0.03	0.69	26	8	18
placed_in	1	0.01	0.67	15	5	10
delays_in	1	0.02	0.71	14	4	10
pediatrician	1	0.02	0.85	13	2	11
been_in_foster	1	0.01	0.80	10	2	8
he_struggles	1	0.01	0.80	10	2	8
Glenrose	1	0.01	0.80	10	2	8
autism_EOL	1	0.01	1.00	7	0	7
he_struggles_with	1	0.01	1.00	6	0	6
stem_from	1	0.01	1.00	6	0	6
child_is_too	1	0.01	1.00	5	0	5
child_presents_with	1	0.01	1.00	5	0	5

Emotional maturity (EMO)

In Table 8 are presented descriptions of children with or without emotional problems. When describing children with emotional maturity problems, they are often described with parenting, attention issues, emotional and behavioural issues, issues related to control, and mention of a pediatrician. Conversely, children who have no emotional maturity problems will likely be confident, eager, described as very well in any manner, and is able to write.

Table 8: Vulnerability based on features of emotional maturity

Feature	Predictor	Kappa	Precision	Hit	Hit 0	Hit 1
curriculum	0	0.01	0.93	85	79	6
reading	0	0.01	0.93	82	76	6
very_well	0	0.01	0.93	70	65	5
eager	0	0.01	0.98	43	42	1
writing	0	0.01	0.93	40	37	3
loves	0	0.01	0.94	33	31	2
confident	0	0.00	0.94	31	29	2
happy	0	0.00	0.93	30	28	2
quickly	0	0.01	0.97	29	28	1
attended_junior	0	0.00	0.96	28	27	1
english_EOL	0	0.00	0.95	21	20	1
aware	0	0.00	0.95	19	18	1
the_english	0	0.00	0.94	18	17	1
arrived	0	0.00	0.94	18	17	1
very_shy_and	0	0.00	0.94	18	17	1
moved_from	0	0.00	0.93	15	14	1
parenting	1	0.01	1.00	9	0	9
autism_EOL	1	0.01	1.00	7	0	7
with_attention	1	0.01	1.00	5	0	5
child_presents_with	1	0.01	1.00	5	0	5
placed_in_foster	1	0.01	1.00	5	0	5
a_pediatrician	1	0.01	0.90	10	1	9
emotional_	1	0.01	0.89	9	1	8
the_help	1	0.01	0.86	7	1	6
autism_and	1	0.01	0.86	7	1	6
behavior_issues	1	0.01	0.86	7	1	6
unstable_home	1	0.01	0.86	7	1	6
at_risk	1	0.01	0.86	7	1	6
doesn't_get	1	0.01	0.86	7	1	6
adhd	1	0.03	0.83	23	4	19
the_glenrose	1	0.01	0.80	10	2	8
to_control	1	0.01	0.80	10	2	8
oneonone	1	0.01	0.80	10	2	8
a_psychologist	1	0.01	0.80	10	2	8

Language & thinking skills (LAN)

Features of comments that describe children with issues on language & thinking skills are: not ready, issues related to being on or with a program, have been assessed, and have needs (Table 9). Conversely, features that most prominently describe children with no vulnerabilities in the area of language and thinking are often described as: being bright, do well in their activities, and also enjoy and like what they do. Surprisingly, moderate descriptors such as *somewhat* or *seem to* do portray children without vulnerabilities in the language & thinking skills domain.

Table 9: Vulnerability based on features of language & thinking skills

Feature	Predictor	Kappa	Precision	Hit	Hit 0	Hit 1
very_bright	0	0.01	1.00	42	42	0
well_in	0	0.01	1.00	38	38	0
quickly	0	0.00	1.00	29	29	0
very_ready	0	0.00	1.00	28	28	0
enjoys	0	0.00	1.00	25	25	0
Montessori	0	0.00	1.00	24	24	0
somewhat	0	0.00	1.00	23	23	0
is_very_bright	0	0.00	1.00	23	23	0
was_ready_for	0	0.00	1.00	21	21	0
doing_very_well	0	0.00	1.00	20	20	0
likes	0	0.00	1.00	20	20	0
is_doing_very	0	0.00	1.00	19	19	0
seem_to	0	0.00	1.00	19	19	0
is_reading	0	0.00	1.00	19	19	0
enthusiastic	0	0.00	1.00	18	18	0
outgoing	0	0.00	1.00	17	17	0
advanced	0	0.00	1.00	17	17	0
not_ready_to	1	0.02	0.80	15	3	12
assess	1	0.02	0.83	12	2	10
of_program_unit	1	0.01	1.00	8	0	8
reserve	1	0.01	1.00	7	0	7
year_of_program	1	0.01	1.00	6	0	6
needs_one	1	0.01	1.00	6	0	6
name_and	1	0.01	1.00	5	0	5
he_is_beginning	1	0.01	1.00	5	0	5
home_on	1	0.01	1.00	5	0	5
may_need	1	0.01	1.00	5	0	5
has_little	1	0.01	1.00	5	0	5
living_in_a	1	0.01	1.00	5	0	5

Communication & general knowledge (COM)

Finally, a summary of communication skills & general knowledge features is presented in Table 10. Children who are new to Canada, have language delays, described as very poor in activities, do not have or not receptive to certain conditions all describe those having communication and general knowledge problems. In describing those with no communication & general knowledge issues, features of the student include being eager, do very well, outgoing, and are simply ready for school.

Table 10: Vulnerability based on features of communication & general knowledge

Feature	Predictor	Kappa	Precision	Hit	Hit 0	Hit 1
curriculum	0	0.01	0.88	85	75	10
eager	0	0.01	0.88	43	38	5
school_readiness	0	0.00	0.88	25	22	3
doing_very_well	0	0.00	1.00	20	20	0
is_doing_very	0	0.00	1.00	19	19	0
outgoing	0	0.00	1.00	17	17	0
advanced	0	0.00	1.00	17	17	0
math_curriculum	0	0.00	1.00	16	16	0
achieving	0	0.00	1.00	16	16	0
canada	1	0.04	0.74	53	14	39
language_delays	1	0.02	0.76	21	5	16
very_poor	1	0.01	0.75	20	5	15
does_not_have	1	0.01	0.75	16	4	12
receptive_and	1	0.01	0.80	15	3	12
not_yet	1	0.01	0.73	15	4	11
of_puf	1	0.01	1.00	10	0	10
lives_on	1	0.01	1.00	8	0	8
3rd	1	0.01	1.00	8	0	8
colony	1	0.01	1.00	8	0	8
autism_EOL	1	0.01	1.00	7	0	7
and_is_in	1	0.01	1.00	7	0	7
and_language_delays	1	0.01	1.00	7	0	7
nonverbal	1	0.01	1.00	7	0	7

Conclusions drawn from NLP analysis

In Part 2, we provided a mechanism to identify features of vulnerable and non-vulnerable children for each domain of the EDI. These features were summarized and highlighted to demonstrate a profile of the children who are lagging behind in any of the five areas of development. Future studies could expand on such analyses of discriminative features and perhaps provide a vignette to the public to describe the warning signs of a specific area of vulnerability. Although quantitative measures provide a good method for aggregating community results as a whole, dissemination of these results may require qualitative descriptors to supplement other validation designs, and thereby enhance the evaluation and usefulness of a tool cross-culturally.

Limitations of this approach include the exclusion of responses that do not provide comments for some children. While we were successful in demonstrating how this analytic approach can be used to identify features that portray vulnerable children, only 9% of the response set offered any qualitative information. Moreover, providing such comments may be dependent on the teachers' own style of commenting or giving feedback where some teachers like to provide qualitative feedback for all children, while others prefer to provide only objective or quantitative information. Therefore, it is likely that we are only capturing a small proportion of teachers who are willing to express their perspectives in writing. Another limitation of this study is that while only those responses that are written in English were analyzed, written comments in French were totally ignored. This may underestimate our classifications, to some extent.

In large scale surveys, the extraction and the reporting of relevant information, especially of qualitative information is a difficult task. A statistical summary that provides a narrow view of the context is often the only viable method. But with the emergence of machine learning techniques, sophisticated methods can now be used to extract meaningful information from large data sources, such as the EDI. With all its limitations, then, Part 2 of the study will have the potential to provide additional insights into vulnerabilities of young children in a community. More specifically, we see the advantage of such designs as NLP in exploring the qualitative information in large scale surveys, viewed through a different lens that have the potential to enhance data analysis, information extraction, and knowledge translation of developmental challenges in children.

4. Conclusions and final considerations

The purpose of the Early Childhood Mapping (ECMap) project is to provide a channel of communication and information about young children's development across different communities. The benchmarking and indexing of information have thus far progressed at a quick pace, where data have been collected across all of Alberta. The project uses the Early Development Instrument (EDI) as an index to provide a common measure across communities and provide an ongoing conversation of development within each community. The methods used and demonstrated in this study seek to enhance this reporting structure to ensure information provided to the community have a positive association and applicability to the general public.

As every community is composed of many diverse groups of people of different cultural and linguistic backgrounds, in the first study we explored whether the instrument elicited response biases for a specific group of the child population. There are evident risks of assessment of young children's communication & general knowledge skills where the pendulum swaying too far from minority children or the focus on EDI is exclusively on children of European or North American backgrounds. From the results of the Part 1 of our study, in relation to communication & general knowledge domain, there are questions as to whether the component can ever be robust enough to inform policy in a cross-cultural context. In Part 2 of the study, we demonstrated how natural language processing techniques can be used to identify features of teachers' comments that are characteristics of vulnerable and non-vulnerable children. This part of the study provided a starting point for the EDI users and researchers in understanding the use of qualitative information to better inform teachers and parents on the risk factors associated with developmental delays in children. Building on the two constituent parts of the study, we have presented a pair of state of the art solutions to enhance the use of the EDI thereby laying a path to improve the reporting mechanisms to the community, especially when assessing members of subpopulations in multicultural societies like Canada.

References

Ball, J. (2010). Enhancing learning of children from diverse language backgrounds: Mother tongue-based bilingual or multilingual education in the early years. Paper commissioned for UNESCO. Retrieved April 27, 2011, from *http://unesdoc.unesco.org/images/0018/001869/186961e.pdf*.

Belov, D., & Knezevich, L. (2008*). Automatic prediction of item difficulty based on semantic measures.* Research Report, 08-04. Law School Admissions Council, Newtown, PA.

Benson, C. (2002). Real and potential benefits of bilingual programmes in developing countries. *International Journal of Bilingual Education and Bilingualism, 5* (6), 303-317.

Butcher, J. N. & Garcia, R. E. (1978). Cross-national application of psychological tests. *Personnel and Guidance Journal, 5*6, 472-75.

Chauvin, Y. & Rumelhart, D. (1995). *Backpropagation: Theory, Architectures, and Applications.* Hillsdale, NJ: Erlbaum.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design & analysis issues from field settings. Boston, MA: Houghton Mifflin Company.

Forer, B., & Zumbo, B. D. (2011). Validation of multilevel constructs: Validation methods and empirical findings for the EDI. *Social Indicators Research*, 103(2), 231-265, DOI: 10.1007/s11205-011-9844-3.

Government of Alberta (2010). Budget 2010: Striking the Right Balance. Government of Alberta Strategic Business Plan and Ministry Business Plans 2010-13.

Guhn, M., Gadermann, A., & Zumbo, B. (2007). Does the EDI Measure School Readiness in the Same Way Across Different Groups of Children? *Early Education & Development,* 18(3), 453-472.

Hovens, M. (2002). Bilingual education in West Africa: Does it work? *International Journal of Bilingual Education and Bilingualism, 5* (5), 249-266.

Hymel, S., Le Mare, L., & McKee, W. (2011). The Early Development Instrument (EDI): An examination of convergent and discriminant validity. *Social Indicators Research*, 103(2), 267-282, DOI: 10.1007/s11205-011-9845-2.

Janus, M. (2006). Early Development Instrument: An indicator of developmental health at school entry. Paper presented at the International Conference on Measuring Early Child Development. Vaudreuil, QC, April.

Janus, M., Brinkman, S., & Duku, E. (2011). Validity and psychometric properties of Early Development Instrument in Canada, Australia, United States, and Jamaica. *Social Indicators Research.* 103(2), 283-297. doi: 10.1007/s11205-011-9846-1.

Jodoin, M., & Gierl, M. (1999). Evaluating Type I error and power using an effect size measure with the logistic regression procedure for DIF detection. Paper presented at the annual meeting of the National Council for Measurement in Education.

Karamanis, N., Ha, L. A. and Mitkov, R. (2006). Generating Multiple-Choice Test Items from Medical Text: A Pilot Study. *In Proceedings of INLG 2006*. Sydney, Australia.

Kosonen, K. (2005). Education in local languages: Policy and practice in Southeast Asia. *First languages first: Community-based literacy programmes for minority language contexts in Asia.* Bangkok: UNESCO Bangkok.

Krishnan, V., Wang, H., Babenko, O., & Lynch, S. (2011). Alberta's Children in Their Early Years of Development: An Analysis of the Early Development Instrument (EDI), 2010. ECMap, Community-University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta, Canada.

Krishnan, V. (2010). A comparison of Principal Components Analysis and Factor Analysis for uncovering the Early Development Instrument (EDI) domains, Early Child Mapping (ECMap) Project Alberta, Community-University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta.

Krishnan, V. (2013). The Early Child Development Instrument (EDI): An item analysis using Classical Test Theory (CTT) on Alberta's data, Early Child Mapping (ECMap) Project Alberta, Community-University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta.

Lai, H., & Becker, K. (2010). Using Artificial Neural Networks to Identify Enemy Item Pairs. Poster presented at the National Council for Measurement in Education. Denver, CO., April.

Mayfield, E. & Rose, C. (2012). LightSIDE: Open Source Machine Learning for Text. Retrieved on 1 May, 2012 from http://www.cs.cmu.edu/~cprose/SIDE.html.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Retrieved on 1 May, 2012 from http://www.R-project.org/.

Santos, R., Brownell, M., Ekuma, O., Mayer, T. & Soodeen, R. (2012). The Early Development Instrument (EDI) in Manitoba: Linking Socioeconomic Adversity and Biological Vulnerability at Birth to Children's Outcomes at Age 5. Winnipeg, MB: Manitoba Centre for Health Policy.

Sireci, S. G. & Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language Testing*, 20(2), 148-166.

Statistics Canada (2008). Ethnic Diversity Survey: Portrait of a multicultural society (Catalogue no. 89-593-X1E). Retrieved March 15, 2013, from http://www.statscan.ca/cgi-bin/downpub/freepub.cgi.

Swaminathan, H., & Rogers, J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. Journal of Educational Measurement. 27(4), 361-370.

The Muttart Foundation (2010). In *the Best Interests of Children* and Families: A Discussion of Early Childhood Education and Care in Alberta. Retrieved August 12, 2012, from www.muttart.org/.../ECEC%20Discussion%20Paper%201011_0.pdf

UNESCO (2007). Strong Foundations: Early Childhood Care and Education. Retrieved March 15, 2009, from http://www.unesco.org/en/education/efareport/reports/2007-early-childhood/

Appendix A: DIF results for all five domains by selected characteristics of children

Component	Group	Item Number	Uniform DIF	Non Uniform DIF	R2		
					(Sub Score Only)	(SubScore +DIF CAT)	(SubScore+DIFCat+ interaction)
COM	Gender	1	0.00	0.02	0.37	0.37	0.38
COM	Gender	2	0.00	0.02	0.37	0.37	0.39
COM	Gender	3	0.00	0.02	0.39	0.39	0.41
COM	Gender	4	0.00	0.02	0.37	0.37	0.40
COM	Gender	5	0.00	0.02	0.37	0.37	0.38
COM	Gender	6	0.00	0.02	0.36	0.36	0.38
COM	Gender	7	0.00	0.02	0.31	0.31	0.33
COM	Gender	8	0.00	0.01	0.27	0.27	0.27
COM	ESL	1	0.07	0.12	0.28	0.35	0.40
COM	ESL	2	0.03	0.07	0.30	0.34	0.37
COM	ESL	3	0.04	0.10	0.31	0.35	0.41
COM	ESL	4	0.01	0.06	0.32	0.33	0.38
COM	ESL	5	0.03	0.09	0.29	0.32	0.39
COM	ESL	6	0.03	0.08	0.29	0.32	0.38
COM	ESL	7	0.00	0.05	0.26	0.27	0.31
COM	ESL	8	0.01	0.04	0.20	0.21	0.24
COM	South Asian	1	0.04	0.11	0.25	0.29	0.36
COM	South Asian	2	0.02	0.04	0.42	0.44	0.47
COM	South Asian	3	0.05	0.15	0.38	0.42	0.53
COM	South Asian	4	0.00	0.03	0.37	0.38	0.40
COM	South Asian	5	0.02	0.10	0.25	0.27	0.35
COM	South Asian	6	0.03	0.09	0.25	0.28	0.33
COM	South Asian	7	0.00	0.10	0.37	0.37	0.46
COM	South Asian	8	0.01	0.03	0.38	0.39	0.40
EMO	Gender	1	0.00	0.00	0.40	0.40	0.40
EMO	Gender	2	0.00	0.00	0.45	0.45	0.45
EMO	Gender	3	0.00	0.00	0.40	0.40	0.41
EMO	Gender	4	0.00	0.00	0.40	0.40	0.40
EMO	Gender	5	0.00	0.01	0.38	0.38	0.38
EMO	Gender	6	0.00	0.00	0.40	0.40	0.40
EMO	Gender	7	0.00	0.00	0.04	0.04	0.04
EMO	Gender	8	0.00	0.00	0.15	0.16	0.17
EMO	Gender	9	0.00	0.01	0.18	0.18	0.18
EMO	Gender	10	0.00	0.02	0.17	0.17	0.18
EMO	Gender	11	0.00	0.00	0.17	0.17	0.17

EMO	Gender	12	0.00	0.01	0.13	0.13	0.13
EMO	Gender	13	0.00	0.00	0.29	0.30	0.31
EMO	Gender	14	0.00	0.00	0.32	0.32	0.33
EMO	Gender	15	0.00	0.01	0.30	0.30	0.31
EMO	Gender	16	0.00	0.01	0.29	0.29	0.30
EMO	Gender	17	0.00	0.01	0.17	0.17	0.18
EMO	Gender	18	0.00	0.01	0.30	0.30	0.31
EMO	Gender	19	0.00	0.01	0.30	0.30	0.31
EMO	Gender	20	0.00	0.01	0.27	0.27	0.29
EMO	Gender	21	0.00	0.02	0.34	0.34	0.34
EMO	Gender	22	0.00	0.01	0.18	0.18	0.18
EMO	Gender	23	0.00	0.00	0.12	0.12	0.13
EMO	Gender	24	0.00	0.00	0.11	0.11	0.11
EMO	Gender	25	0.01	0.01	0.11	0.11	0.11
EMO	Gender	26	0.00	0.00	0.12	0.12	0.12
EMO	Gender	27	0.00	0.00	0.19	0.19	0.19
EMO	Gender	28	0.00	0.00	0.02	0.04	0.04
EMO	Gender	29	0.00	0.00	0.35	0.35	0.35
EMO	Gender	30	0.01	0.02	0.35	0.35	0.35
EMO	ESL	1	0.00	0.01	0.36	0.36	0.37
EMO	ESL	2	0.00	0.00	0.39	0.39	0.39
EMO	ESL	3	0.00	0.01	0.41	0.41	0.42
EMO	ESL	4	0.00	0.01	0.44	0.44	0.45
EMO	ESL	5	0.00	0.01	0.39	0.40	0.40
EMO	ESL	6	0.00	0.01	0.40	0.40	0.40
EMO	ESL	7	0.00	0.01	0.38	0.39	0.39
EMO	ESL	8	0.00	0.01	0.43	0.43	0.44
EMO	ESL	9	0.00	0.00	0.03	0.04	0.04
EMO	ESL	10	0.00	0.00	0.14	0.14	0.14
EMO	ESL	11	0.00	0.00	0.14	0.15	0.15
EMO	ESL	12	0.00	0.00	0.16	0.16	0.16
EMO	ESL	13	0.00	0.00	0.14	0.14	0.14
EMO	ESL	14	0.00	0.00	0.13	0.13	0.13
EMO	ESL	15	0.00	0.00	0.30	0.30	0.30
EMO	ESL	16	0.00	0.00	0.32	0.32	0.32
EMO	ESL	17	0.00	0.00	0.30	0.30	0.30
EMO	ESL	18	0.00	0.00	0.27	0.27	0.27
EMO	ESL	19	0.00	0.01	0.15	0.16	0.17
EMO	ESL	20	0.00	0.01	0.26	0.26	0.26
EMO	ESL	21	0.00	0.00	0.26	0.26	0.26

EMO	ESL	22	0.00	0.00	0.27	0.27	0.27
EMO	ESL	23	0.00	0.00	0.33	0.33	0.33
EMO	ESL	24	0.00	0.01	0.16	0.17	0.17
EMO	ESL	25	0.00	0.01	0.11	0.11	0.11
EMO	ESL	26	0.01	0.01	0.09	0.10	0.10
EMO	ESL	27	0.00	0.01	0.10	0.10	0.11
EMO	ESL	28	0.00	0.01	0.09	0.10	0.10
EMO	ESL	29	0.00	0.00	0.17	0.17	0.17
EMO	ESL	30	0.00	0.00	0.03	0.03	0.03
EMO	South Asian	1	0.01	0.02	0.39	0.40	0.41
EMO	South Asian	2	0.00	0.00	0.40	0.40	0.40
EMO	South Asian	3	0.00	0.00	0.44	0.44	0.44
EMO	South Asian	4	0.00	0.01	0.45	0.46	0.46
EMO	South Asian	5	0.01	0.02	0.42	0.43	0.44
EMO	South Asian	6	0.00	0.01	0.40	0.40	0.41
EMO	South Asian	7	0.01	0.02	0.39	0.40	0.41
EMO	South Asian	8	0.00	0.01	0.45	0.45	0.46
EMO	South Asian	9	0.00	0.00	0.02	0.02	0.02
EMO	South Asian	10	0.00	0.00	0.15	0.15	0.15
EMO	South Asian	11	0.00	0.00	0.14	0.14	0.14
EMO	South Asian	12	0.00	0.00	0.15	0.15	0.15
EMO	South Asian	13	-0.01	0.00	0.10	0.09	0.09
EMO	South Asian	14	0.00	0.00	0.12	0.12	0.12
EMO	South Asian	15	0.01	0.02	0.27	0.28	0.29
EMO	South Asian	16	0.01	0.02	0.31	0.32	0.33
EMO	South Asian	17	0.01	0.02	0.25	0.26	0.27
EMO	South Asian	18	0.00	0.00	0.25	0.25	0.25
EMO	South Asian	19	0.01	0.04	0.15	0.16	0.19
EMO	South Asian	20	0.01	0.02	0.27	0.28	0.29
EMO	South Asian	21	0.01	0.01	0.24	0.25	0.25
EMO	South Asian	22	0.00	0.00	0.24	0.24	0.25
EMO	South Asian	23	0.01	0.01	0.33	0.34	0.34
EMO	South Asian	24	0.00	0.01	0.13	0.13	0.14
EMO	South Asian	25	0.00	0.00	0.12	0.12	0.12
EMO	South Asian	26	0.00	0.00	0.09	0.09	0.09
EMO	South Asian	27	0.00	0.01	0.08	0.08	0.08
EMO	South Asian	28	0.00	0.00	0.11	0.10	0.11
EMO	South Asian	29	0.00	0.00	0.16	0.16	0.16
EMO	South Asian	30	0.00	0.01	0.03	0.03	0.04
LAN	ESL	1	-0.02	-0.02	0.01	-0.02	-0.01

LAN	ESL	2	-0.01	-0.01	0.09	0.08	0.08
LAN	ESL	3	0.00	0.00	0.21	0.20	0.20
LAN	ESL	4	0.00	0.00	0.26	0.26	0.26
LAN	ESL	5	0.00	0.00	0.36	0.36	0.36
LAN	ESL	6	0.01	0.03	0.36	0.37	0.39
LAN	ESL	7	0.00	0.00	0.24	0.24	0.24
LAN	ESL	8	0.00	0.00	0.44	0.44	0.44
LAN	ESL	9	0.00	0.00	0.58	0.58	0.58
LAN	ESL	10	0.00	0.00	0.47	0.47	0.47
LAN	ESL	11	0.00	0.00	0.10	0.10	0.10
LAN	ESL	12	0.00	0.00	0.19	0.19	0.19
LAN	ESL	13	0.00	0.00	0.26	0.26	0.26
LAN	ESL	14	-0.01	-0.01	0.10	0.09	0.09
LAN	ESL	15	0.00	0.01	0.26	0.26	0.26
LAN	ESL	16	0.00	0.00	0.30	0.30	0.30
LAN	ESL	17	0.00	0.00	0.27	0.27	0.27
LAN	ESL	18	0.00	0.00	0.22	0.22	0.22
LAN	ESL	19	0.00	0.00	0.20	0.20	0.20
LAN	ESL	20	0.00	0.00	0.16	0.16	0.16
LAN	ESL	21	0.00	0.00	0.23	0.22	0.22
LAN	ESL	22	0.00	0.00	0.31	0.31	0.31
LAN	ESL	23	0.00	0.00	0.28	0.28	0.28
LAN	ESL	24	0.00	0.00	0.28	0.27	0.27
LAN	ESL	25	0.00	0.02	0.17	0.17	0.19
LAN	ESL	26	0.00	0.02	0.17	0.17	0.19
LAN	Gender	1	-0.03	-0.01	0.00	-0.03	-0.01
LAN	Gender	2	0.00	0.01	0.08	0.08	0.09
LAN	Gender	3	0.00	0.01	0.19	0.19	0.20
LAN	Gender	4	0.00	0.00	0.28	0.28	0.28
LAN	Gender	5	0.00	0.00	0.35	0.35	0.35
LAN	Gender	6	0.00	0.00	0.33	0.33	0.33
LAN	Gender	7	0.00	0.00	0.23	0.22	0.22
LAN	Gender	8	0.00	0.00	0.45	0.44	0.45
LAN	Gender	9	0.00	0.01	0.60	0.60	0.61
LAN	Gender	10	0.00	0.00	0.49	0.49	0.49
LAN	Gender	11	0.00	0.02	0.11	0.11	0.13
LAN	Gender	12	0.00	0.00	0.18	0.18	0.19
LAN	Gender	13	0.03	0.05	0.27	0.30	0.33
LAN	Gender	14	-0.01	0.00	0.11	0.10	0.11
LAN	Gender	15	0.00	0.00	0.23	0.23	0.23

LAN	Gender	16	0.00	0.00	0.31	0.31	0.31
LAN	Gender	17	0.00	0.00	0.29	0.29	0.29
LAN	Gender	18	0.00	0.00	0.22	0.22	0.22
LAN	Gender	19	0.00	0.00	0.18	0.18	0.18
LAN	Gender	20	0.00	0.00	0.16	0.16	0.16
LAN	Gender	21	0.00	0.00	0.22	0.22	0.22
LAN	Gender	22	0.00	0.00	0.31	0.31	0.31
LAN	Gender	23	0.00	0.00	0.28	0.28	0.29
LAN	Gender	24	0.00	0.00	0.28	0.28	0.28
LAN	Gender	25	-0.01	0.00	0.14	0.13	0.13
LAN	Gender	26	0.00	0.00	0.17	0.17	0.17
LAN	South Asian	1	-0.05	-0.05	-0.07	-0.13	-0.12
LAN	South Asian	2	-0.02	-0.01	0.06	0.04	0.05
LAN	South Asian	3	0.00	0.00	0.18	0.17	0.18
LAN	South Asian	4	0.00	0.01	0.18	0.18	0.19
LAN	South Asian	5	0.00	0.00	0.35	0.34	0.34
LAN	South Asian	6	0.01	0.03	0.37	0.38	0.39
LAN	South Asian	7	-0.01	-0.01	0.21	0.21	0.21
LAN	South Asian	8	0.00	0.00	0.43	0.42	0.42
LAN	South Asian	9	0.00	0.00	0.60	0.60	0.60
LAN	South Asian	10	0.00	0.00	0.45	0.45	0.45
LAN	South Asian	11	-0.01	-0.01	0.09	0.08	0.08
LAN	South Asian	12	-0.01	-0.01	0.17	0.17	0.17
LAN	South Asian	13	0.00	0.00	0.29	0.29	0.29
LAN	South Asian	14	-0.02	-0.02	0.05	0.03	0.03
LAN	South Asian	15	0.00	0.00	0.27	0.27	0.27
LAN	South Asian	16	0.00	0.00	0.29	0.29	0.29
LAN	South Asian	17	0.00	0.00	0.32	0.32	0.32
LAN	South Asian	18	-0.01	0.00	0.21	0.20	0.20
LAN	South Asian	19	-0.01	0.00	0.17	0.17	0.17
LAN	South Asian	20	-0.01	0.00	0.13	0.12	0.12
LAN	South Asian	21	-0.01	-0.01	0.23	0.22	0.23
LAN	South Asian	22	0.00	0.00	0.30	0.30	0.30
LAN	South Asian	23	0.00	0.01	0.26	0.26	0.27
LAN	South Asian	24	0.00	0.00	0.26	0.26	0.26
LAN	South Asian	25	-0.01	0.00	0.15	0.14	0.14
LAN	South Asian	26	0.00	0.02	0.19	0.20	0.21
PHY	Gender	1	0.00	0.00	0.11	0.12	0.12
PHY	Gender	2	0.00	0.00	0.16	0.16	0.16
PHY	Gender	3	0.00	0.00	0.10	0.10	0.10

PHY	Gender	4	0.00	0.00	0.01	0.01	0.01
PHY	Gender	5	0.00	0.00	0.03	0.03	0.03
PHY	Gender	6	0.00	0.00	0.03	0.03	0.04
PHY	Gender	7	0.00	0.01	0.17	0.17	0.18
PHY	Gender	8	0.01	0.02	0.42	0.43	0.44
PHY	Gender	9	0.00	0.01	0.49	0.49	0.50
PHY	Gender	10	0.00	0.00	0.04	0.04	0.04
PHY	Gender	11	0.00	0.00	0.45	0.45	0.45
PHY	Gender	12	0.00	0.00	0.53	0.53	0.53
PHY	Gender	13	0.00	0.00	0.00	0.00	0.00
PHY	ESL	1	0.00	0.01	0.10	0.10	0.11
PHY	ESL	2	0.00	0.01	0.13	0.13	0.13
PHY	ESL	3	0.00	0.00	0.12	0.12	0.12
PHY	ESL	4	0.00	0.00	0.01	0.01	0.01
PHY	ESL	5	0.00	0.00	0.03	0.03	0.03
PHY	ESL	6	0.00	0.00	0.02	0.02	0.02
PHY	ESL	7	0.00	0.00	0.17	0.17	0.17
PHY	ESL	8	0.00	0.00	0.42	0.42	0.42
PHY	ESL	9	0.00	0.00	0.49	0.49	0.49
PHY	ESL	10	0.00	0.00	0.03	0.03	0.04
PHY	ESL	11	0.00	0.00	0.49	0.49	0.49
PHY	ESL	12	0.00	0.00	0.55	0.55	0.55
PHY	ESL	13	0.00	0.00	0.00	0.00	0.00
PHY	South Asian	1	0.00	0.00	0.10	0.10	0.10
PHY	South Asian	2	0.00	0.01	0.13	0.14	0.15
PHY	South Asian	3	0.00	0.00	0.15	0.15	0.15
PHY	South Asian	4	0.01	0.05	0.00	0.01	0.05
PHY	South Asian	5	-0.01	-0.01	0.02	0.01	0.01
PHY	South Asian	6	0.00	0.00	0.02	0.02	0.02
PHY	South Asian	7	0.00	0.01	0.11	0.11	0.12
PHY	South Asian	8	0.00	0.00	0.41	0.41	0.42
PHY	South Asian	9	0.00	0.00	0.46	0.46	0.46
PHY	South Asian	10	0.00	0.00	0.02	0.02	0.02
PHY	South Asian	11	0.00	0.00	0.49	0.49	0.49
PHY	South Asian	12	0.00	0.00	0.53	0.53	0.53
PHY	South Asian	13	0.00	0.00	0.00	0.00	0.00
SOC	Gender	1	0.00	0.00	0.47	0.47	0.47
SOC	Gender	2	0.00	0.00	0.47	0.47	0.47
SOC	Gender	3	0.00	0.00	0.52	0.52	0.52
SOC	Gender	4	0.00	0.00	0.43	0.43	0.43

SOC	Gender	5	0.00	0.00	0.55	0.55	0.55
SOC	Gender	6	0.00	0.00	0.43	0.43	0.44
SOC	Gender	7	0.00	0.00	0.48	0.48	0.48
SOC	Gender	8	0.00	0.00	0.30	0.30	0.30
SOC	Gender	9	0.00	0.00	0.41	0.40	0.41
SOC	Gender	10	0.00	0.00	0.43	0.43	0.43
SOC	Gender	11	0.00	0.00	0.48	0.47	0.48
SOC	Gender	12	0.00	0.00	0.50	0.50	0.50
SOC	Gender	13	0.00	0.00	0.55	0.55	0.55
SOC	Gender	14	0.00	0.00	0.41	0.41	0.41
SOC	Gender	15	0.00	0.00	0.48	0.48	0.48
SOC	Gender	16	0.00	0.01	0.43	0.42	0.43
SOC	Gender	17	0.00	0.00	0.39	0.39	0.39
SOC	Gender	18	0.00	0.00	0.29	0.29	0.29
SOC	Gender	19	0.00	0.00	0.17	0.17	0.17
SOC	Gender	20	0.00	0.00	0.21	0.21	0.21
SOC	Gender	21	0.00	0.00	0.25	0.25	0.25
SOC	Gender	22	0.00	0.00	0.45	0.45	0.45
SOC	Gender	23	0.00	0.00	0.41	0.41	0.41
SOC	Gender	24	0.00	0.00	0.49	0.49	0.49
SOC	Gender	25	0.00	0.00	0.43	0.43	0.43
SOC	Gender	26	0.00	0.00	0.35	0.34	0.35
SOC	South Asian	1	0.00	0.00	0.44	0.44	0.44
SOC	South Asian	2	-0.01	-0.01	0.44	0.44	0.44
SOC	South Asian	3	-0.01	-0.01	0.48	0.48	0.48
SOC	South Asian	4	-0.01	-0.01	0.41	0.41	0.41
SOC	South Asian	5	-0.01	-0.01	0.50	0.49	0.49
SOC	South Asian	6	-0.01	-0.01	0.38	0.37	0.37
SOC	South Asian	7	0.00	0.00	0.40	0.40	0.41
SOC	South Asian	8	-0.01	0.00	0.34	0.33	0.34
SOC	South Asian	9	-0.02	0.00	0.31	0.29	0.30
SOC	South Asian	10	-0.01	0.00	0.38	0.37	0.38
SOC	South Asian	11	-0.01	0.00	0.42	0.42	0.42
SOC	South Asian	12	-0.01	-0.01	0.46	0.46	0.46
SOC	South Asian	13	-0.01	-0.01	0.53	0.52	0.52
SOC	South Asian	14	-0.01	-0.01	0.37	0.37	0.37
SOC	South Asian	15	-0.01	-0.01	0.42	0.41	0.41
SOC	South Asian	16	-0.01	0.01	0.39	0.38	0.40
SOC	South Asian	17	0.00	0.00	0.35	0.35	0.35
SOC	South Asian	18	-0.01	0.01	0.24	0.24	0.26

SOC	South Asian	19	-0.02	-0.02	0.11	0.09	0.09
SOC	South Asian	20	-0.02	-0.02	0.15	0.14	0.14
SOC	South Asian	21	-0.01	-0.01	0.25	0.23	0.24
SOC	South Asian	22	0.00	0.01	0.44	0.45	0.45
SOC	South Asian	23	-0.01	-0.01	0.39	0.38	0.38
SOC	South Asian	24	-0.01	-0.01	0.48	0.48	0.48
SOC	South Asian	25	-0.01	-0.01	0.43	0.42	0.42
SOC	South Asian	26	0.00	0.00	0.25	0.25	0.25
SOC	ESL	1	0.00	0.00	0.45	0.45	0.46
SOC	ESL	2	0.00	0.00	0.46	0.46	0.46
SOC	ESL	3	0.00	0.00	0.51	0.51	0.51
SOC	ESL	4	0.00	0.00	0.43	0.43	0.43
SOC	ESL	5	0.00	0.00	0.54	0.54	0.54
SOC	ESL	6	0.00	0.00	0.41	0.41	0.41
SOC	ESL	7	0.00	0.00	0.45	0.44	0.45
SOC	ESL	8	0.00	0.00	0.31	0.31	0.31
SOC	ESL	9	0.00	0.00	0.38	0.38	0.38
SOC	ESL	10	0.00	0.00	0.43	0.42	0.43
SOC	ESL	11	0.00	0.00	0.44	0.44	0.44
SOC	ESL	12	0.00	0.00	0.49	0.49	0.49
SOC	ESL	13	0.00	0.00	0.55	0.54	0.55
SOC	ESL	14	0.00	0.00	0.40	0.40	0.40
SOC	ESL	15	0.00	0.00	0.45	0.45	0.45
SOC	ESL	16	0.00	0.00	0.44	0.44	0.44
SOC	ESL	17	0.00	0.00	0.36	0.37	0.37
SOC	ESL	18	0.00	0.00	0.28	0.28	0.28
SOC	ESL	19	0.00	0.00	0.18	0.18	0.18
SOC	ESL	20	0.00	0.00	0.22	0.22	0.22
SOC	ESL	21	0.00	0.00	0.29	0.28	0.28
SOC	ESL	22	0.00	0.00	0.46	0.46	0.46
SOC	ESL	23	0.00	0.00	0.43	0.42	0.43
SOC	ESL	24	0.00	0.00	0.48	0.48	0.48
SOC	ESL	25	0.00	0.00	0.45	0.44	0.44
SOC	ESL	26	0.00	0.00	0.32	0.32	0.32