"Pick Me, Pick Me, I Want to Be a Counsellor": Assessment of a MEd-Counselling Application Selection Process using Rasch Analysis and Generalizability Theory

by

Stefanie Sebok

B.A., University of Victoria, 2008

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF EDUCATION IN COUNSELLING

THE UNIVERSITY OF NORTHERN BRITISH COLUMBIA

July 2010

© Stefanie Sebok, 2010

*

Library and Archives Canada

Published Heritage Branch

395 Wellington Street Ottawa ON K1A 0N4 Canada Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington Ottawa ON K1A 0N4 Canada

> Your file Votre référence ISBN: 978-0-494-75128-2 Our file Notre référence ISBN: 978-0-494-75128-2

NOTICE:

The author has granted a nonexclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or noncommercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission. AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.



Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Abstract

The purpose of this research project was to evaluate the effectiveness of the Many-Facet Rasch Model and Generalizability Theory as applied to the application selection committee for the Masters of Education in Counselling Program at UNBC. These two models investigated the items used to score applicants and assessed the rater characteristics of each member on the application selection committee. This evaluation was used to inform the School of Education and provide feedback to refine the selection process in the future. Overall, the applicant selection process at UNBC produced a unitary score that can be used to rank all individuals applying to the counselling program. The 5-point rating scale used to evaluate applicants served as an appropriate measurement tool for assessing applicants. The raters who participated as members on the selection committee were fitting both as groups and as individuals in selecting applicants for the counselling program. To conclude, the Many-Facet Rasch Model and Generalizability Theory served as appropriate measurement tools for describing the details of items, raters, and applicants.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	vii
List of figures	viii
Dedication	ix
Acknowledgements	x
Chapter One: Introduction	1
Rationale For the Study	1
Statement of the Problem	2
Theoretical Framework	4
Rasch Model	4
Generalizability Theory	4
Research Question	5
Definition of Terms	5
Chapter Two: Review of Literature	9
Rasch Measurement	9
Dichotomous Rasch Model	11
Rasch-Andrich Rating Scale Model	13
Rasch-Masters Partial Credit Model	14
Many-Facet Rasch Model	15
The Effect of a Rater's Influence	16
Leniency/Severity effect (or generosity error)	16

Central tendency effect	17
Restriction-of-range effect	17
Halo effect	18
Generalizability Theory	19
Relative and absolute decisions	20
Random and fixed facets	20
Decision studies	21
Rasch Meets Generalizability Theory	21
Chapter Three: Research Methods	24
Participants	25
Raters	25
Applicant pool	25
Issues of access to applicant pool	26
Ethical Considerations	26
Instruments	27
Procedures	28
Measures used for Analyzing Data	29
Rasch analysis	29
Generalizability analysis	29
Chapter Four: Results	31
Many-Facet Rasch Analysis	31
Applicant pool	34
Items	36

iv

Raters	37
Applicants	38
Generalizability Analysis	40
Variance components for items	41
Variance components for raters	41
Residual variance component	42
G-Facets analysis	42
Decision studies	45
Part-time Applicants Revisited	48
Rasch revisited	49
Full-time and Part-time Applicants Revisited	51
Probability	55
G-Theory revisited	56
Chapter Five: Discussion and Conclusions	58
Items	59
Raters	60
Applicants	63
Student Raters	64
Conclusions	65
Limitations of the Design	67
Recommendations for the Application Selection Committee	68
Faculty agreement	68
Northern perspective	68

v

Interviews	68
Applicant waitlist	69
Good measurement practice	69
Recommendations for Future Research	70
References	72
Appendix A	76
Appendix B	77
Appendix C	78
Appendix D	79
Appendix E	81
Appendix F	83

٠

vi

List of Tables

Table 1	Items Measurement Report	33
Table 2	Full-time Applicants Items Measurement Report	34
Table 3	Part-time Applicants Items Measurement Report	35
Table 4	Rater Measurement Report	38
Table 5	Applicant Measurement Report	39
Table 6	Estimated G-Study Variance Components	41
Table 7	G-Facets Analysis	43
Table 8	Estimated G-Study Variance Components for Full-Time Applicants	44
Table 9	Estimated G-Study Variance Components for Part-Time Applicants	44
Table 10	D-Studies for Raters	46
Table 11	D-Studies for Items	46
Table 12	Estimated G-Study Variance Components for Combined Part-Time Applicants	49
Table 13	Combined Part-time Applicants Items Measurement Report	50
Table 14	Combined Part-time Applicants Rater Measurement Report	51
Table 15	Combined Prince George and Northwest Applicants Rater Measurement Report	54
Table 16	Estimated G-Study Variance Components for Combined Prince George and Northwest Applicants	57

List of Figures

Figure 1	Wright variable map for relationships among facets for Prince George applicants	32
Figure 2	Alternative D-Studies for determining the optimal number of raters and items	47
Figure 3	Wright variable map for relationships among facets for Prince George and Northwest applicants	53
Figure 4	Probability curves for the 5-point rating scale used to evaluate MEd- Counselling applicants from both the Prince George and Northwest Campuses	55
Figure 5	Scale structure for applicants from Prince George and Northwest campuses	56

For my mom who provided the encouragement, the belief, and the love to get me here.

Acknowledgements

Peter "the Great". Thank you for being the best supervisor, colleague and friend that a person could ever ask for. Being your student has been the most statisfying experience and the memories we share will stay with me forever. Continue to be the great educator you are by sharing your visions and bringing out the best qualities in the students you teach.

Linda. Thank you for all the time you invested in my thesis project. I really appreciate that you were able to be there with me in Whitehorse for my thesis defence.

Kenneth. Thank you for being a part of my thesis project. I found your questions at the defence to be both challenging and insightful.

Robert. Thank you for supporting my thesis work and for sharing your knowledge and expertise of Rasch. I am very honoured to say you were my external examiner.

I would like to acknowledge the School of Education at the University of Northern British Columbia for allowing me to access the data used in this analysis.

John. I appreciate how supportive you were throughout my time in the Masters of Education program. Your calm demeanour and insightful words are great examples of how professors can encourage their students.

Serena. You were my partner in crime throughout this Masters of Education program. Thank you for being there whenever I needed you. You are a very gifted counsellor and a truly wonderful friend.

To my amazing and talented editors: Jeff, Dee, Brandon, and Michael. I appreciate all the time each of you spent to help make my work a masterpiece.

Finally, I want to thank my family and friends who have always been there for me and who continue to encourage me to reach for the stars.

Chapter One: Introduction

Rationale for the Study

People all over the world make decisions about which individuals should be promoted at a job, which individuals are most in need of financial assistance or medical attention, and which individuals are outperforming others in a classroom setting. Each day, people make judgements and decisions based on some kind of formal or informal assessment criteria. When people are making judgements, there is often a bias, whether identified or not, that influences how each individual personally views and interprets the dynamics of a situation. Johnson and Johnson (2003) suggest that when people work together as a group there is more opportunity for the bias that influences their decisions to be identified and addressed. For this particular reason, high stakes decision making usually involves a group approach rather than an individual approach.

Within the academic setting, a group approach to decision making is often used to evaluate the quality and content of work that individuals are trying to research and publish. Groups are also used to assess the effectiveness of particular programs, courses, and instructors. Committees consisting of a variety of individuals with a wealth of experience are constructed so that the operations of a university can be carried out with confidence and ease. Through the use of committees, university policies can be changed or implemented, instructors can be hired or fired, and prospective students can be offered an opportunity to study at an institution.

The purpose of this research study was to assess the overall effectiveness of the MEd-Counselling applicant selection process as it currently exists at the University of Northern British Columbia (UNBC) through an analysis of the effectiveness of the items used by the

1

application selection committee to score applicants and rater characteristics of each of the group members on the application selection committee. This evaluation will inform the School of Education and provide feedback that could be used to further enhance and refine the quality of its application selection process in the future.

Statement of the Problem

Since UNBC first started admitting counselling students, there has been no formal assessment of the overall application process. There has been no attempt to collect empirical evidence to support the validity of the instruments and rating scales used by the selection committee to assess applicants who apply to the counselling program. Therefore, the question of whether the applicant selection process at UNBC produces a unitary score based on the rating scales that can be used to rank all individuals applying to the counselling program remains unanswered. Furthermore, there has never been any formal assessment of the 5-point rating scale used to evaluate the applicants. Without any further investigation, it is indeterminate whether the items' levels of difficulty are appropriately matched to the population of applicants that apply to the MEd-Counselling program each year. In the present study, given that item characteristics were assessed, it logically followed that rater variation must also be addressed (Smith, E. V., 2004). The Many-Facet Rasch Model (Linacre, 1989) can be used to address the query of whether the raters on the selection committee are behaving in a way that demonstrates their ratings of the individuals applying the counselling program fit the model. There is also the question of how the student raters behave in comparison to the faculty raters. If the Rasch model is going to be employed to investigate the rating behaviour of the participants in this study, then further assessment of the Rasch model needs to be conducted to see if it is a viable method of identifying and

compensating for rater differences. The evaluation of the Many-Facet Rasch Model (Linacre, 1989) will be conducted through a comparison with Generalizability Theory, another measurement methodology that has been shown by the literature to investigate applicant data, item characteristics, and rater behaviour.

As UNBC continues to develop and establish itself as a top educational institution, the process of obtaining a seat is going to become more competitive, especially in graduate programs where there are usually a limited number of seats available. Often there are more than enough suitable candidates for these programs; therefore, the selection process becomes about deciding which applicants have the strongest qualities and would be the best fit for the program. Consequently, the selection committee faces added pressure to carefully determine who should be offered a seat in the program. Demonstrated reliability of the instruments and scales used to evaluate applicants on the pre-admission criteria would reassure members of the selection committee, and thus, allow them to better perform as raters.

UNBC was fortunate enough to use a fully crossed design for rating applicants. This means that every member of the application selection committee was responsible for evaluating all applicants on all of the items in the pre-admissions criteria. In situations where a fully crossed design might not have been feasible, information about alternative measures, that could be used to adjust for the variability and bias that would exist, would have been convenient.

In conclusion, evaluating the procedure that was used to select individuals for a competitive program was worthwhile because it provided evidence that UNBC, as an academic institution, is doing everything possible to ensure that the best-suited applicants are being granted letters of admission.

Theoretical Framework

Rasch Model. The Danish mathematician Georg Rasch (1901-1980) created the first Rasch model that researchers in the field of measurement use today. In the 1950s, he was asked to analyze data that was collected from children for the purposes of testing intelligence. Rasch decided to use the multiplicative Poisson model because he felt that it would suitably fit the data (Rasch, 1980). He later applied the Poisson model to measure other data sets. Rasch was able to make a mathematical connection between statistical probabilities and objective measurement, which lead him to develop his own model which uses log-odds transformations and implements "additive" measurement (Linacre, 2010). For further clarification about the fundamentals of additive measurement see Appendix A.

The Rasch model employs the principles of interval measurement to objectively measure the data by taking raw scores of ordinal nature and performing a series of logarithmic transformations to produce data that supports linearity (Linacre, 2010). According to Linacre (2010), if the data does not fit the Rasch model, then another model should be used. This is a revolutionary way of thinking because it gives priority to the model rather than to the data.

Generalizability Theory. In 1972, Cronbach, Glaser, Nanda, and Rajaratnam first introduced the concepts of Generalizability Theory (G-Theory) by extending the work previously done by Hoyt in 1941 (Kieffer, 1999). Using traditional ANOVA methods, G-Theory evaluates multiple sources of measurement variance separately through the process of one single analysis (Atilgan, 2008). In G-Theory, "the object of measurement cannot be a facet" (Kieffer, 1999, p.161) because facets are defined as measures that create unknown error variance. Any sources of variance that do not come from the individuals themselves can be considered a facet. G-Theory looks at how each facet individually contributes to variation in the measurement of a person's overall score in order to obtain a better account of the person's true ability, and thus, makes inferences that can be generalized back to the population. Not only does G-Theory look at these facets individually, but it also looks at the interactions between each facet using an analysis of variance (Shavelson & Webb, 1991).

Research Questions

The purpose of this study was to answer the following questions:

1. What are the strengths and weaknesses of the Rasch Model and Generalizability Theory when applied to an application selection process?

2. Does the applicant selection process at UNBC produce a unitary score based on the rating scales that can be used to rank all individuals applying to the counselling program?

3. Is the 5-point rating scale used to evaluate applicants an appropriate measurement tool?

4. What is the rating behaviour of the raters that participated in selecting applicants for the

MEd-Counselling program both as individuals and as groups?

5. Is the Rasch model a viable method for dealing with rater differences?

Definition of Terms

Many of the terms that will be used throughout this study have specific meaning as they are applied to the Rasch model. The following definitions were taken from the book, *Applying the Rasch model: Fundamental measurement in the human sciences* by Bond and Fox (2007).

<u>Ability Estimate</u>: The location of a person on a variable, inferred by using the collected observations. In this study it would be the applicant's raw score on all of the evaluation items.

<u>Calibration</u>: The procedure of estimating person ability or item difficulty by converting raw scores to logits on an objective measurement scale.

<u>DIF (Differential Item Functioning)</u>: The loss of invariance of item estimates across testing occasions, such as items functioning differently for Prince George applicants and Terrace applicants or differently for males and females. DIF is evidence of item bias.

<u>Facet</u>: An aspect of the measurement condition. In Rasch measurement, the three facets are person ability, item difficulty, and rater severity. In Generalizability Theory, the two facets are item difficulty and rater severity; person ability is not a facet as it is considered the object of measurement.

<u>Infit mean square</u>: One of the two alternative measures that indicate the degree of fit of an item or a person (the other being standardized infit). Infit mean square is a transformation of the residuals, the difference between the predicted and the observed, for easy interpretation. Its expected value is 1. As a general rule, values between 0.70 and 1.30 are regarded as acceptable. Values greater than 1.30 are labelled as "misfitting" and those less than 0.70 as "overfitting."

<u>Item difficulty</u>: The level of resistance to successful performance of the object of measurement on the latent variable. An item with a high level of difficulty should produce a low marginal score. An estimate of an item's underlying difficulty is calculated from the total number of persons in an appropriate sample who succeeded on that item. <u>Latent trait</u>: A characteristic or attribute of a person that can be inferred from the observations of the person's behaviours. <u>Logit</u>: The unit of measurement that results when the Rasch model is used to transform raw scores obtained from ordinal data to log odds ratios on a common interval scale. The value of 0.0 logits is routinely allocated to the mean of the item difficulty estimates.

<u>Many-facets Model</u>: In this model, a version of the Rasch model developed in the work of J. M. Linacre (1989), facets of testing situation in addition to person ability and item difficulty are estimated. Rater, test, or candidate characteristics are often-estimated facets.

<u>Missing data</u>: Data to which there are non-responses for items. Typically, these are items that an applicant did not answer (in this case did not submit), items that were not administered to the applicant, or items that were not judged by a rater.

<u>Noise:</u> Randomness in the data as suggested by the Rasch model or excessive unpredictability in the data, perhaps due to excessive randomness or multidimensionality. <u>Outfit mean square</u>: The measure of degree of fit that is sensitive to outliers, unexpectedly correct responses on hard items or unexpectedly incorrect responses on easy items. Generally, values between 0.60 and 1.40 are generally regarded as acceptable. <u>Raters</u>: Faculty and students who evaluate candidates' test performances in terms of performance criteria.

<u>Residual</u>: The residual values represent the difference between the Rasch model's theoretical expectations and the actual performance.

<u>Unidimensionality</u>: A basic concept in scientific measurement where one attribute of an object is measured at a time. The Rasch model requires a single construct to be underlying the items that form a hierarchical continuum.

Part of the problem with having multiple raters assess and evaluate applicants is the ambiguity that is involved in the process. Rating and scoring the performance of individuals is a difficult task because presenting individuals with a rating scale and asking them to use it in the same way that they would use measuring cups is unrealistic. For this reason, existing rating scales need to be evaluated to see if they are appropriate measurement tools for what the institution is hoping to measure in potential applicants. Examining the rating behaviour of those individuals involved in the process is the other half of evaluating rating scales because information about a rating scale's effectiveness comes from how individuals are interpreting the scale. The ultimate goal of this study was to look at the applicants, items, and rater behaviour to demonstrate that the School of Education at UNBC is engaging in an application selection process that results in selecting the top-ranking applicants who have the background, knowledge, and experience to become good counsellors in the future.

Chapter Two: Review of the Literature

Measurement is assessing and recording observations of things that happen all around us. Everybody uses measurement at some point in their lives. Measurement is observing how fast a car is going, how much of a particular medication is being administered, or how many onions should be added to make the perfect spaghetti sauce. Measurement occurs when a person is concerned with the outcome of what they are doing. Many people will use methods of measurement to provide security and structure to their lives. Measuring something makes it credible and important, as people often measure things that matter in some way or another. Those who are interested in applied forms of measurement know that there are many different types of measurement models available, including factor analysis, general linear models, regression, item response theory, and psychometrics. Often, the challenge with measurement is finding the measurement model that fits with what it is that an individual is interested in measuring. In this study, the researcher was interested in measuring human performance, rating behaviour, and item difficulty and fit. Using a model that employed a multi-facet approach seemed most appropriate to use for the intent of this research. After investigating different measurement models that could be used to answer the research questions put forth by the researcher involved in this study, it appeared that Generalizability Theory and the Many-Facet Rasch Model were the most appropriate.

Rasch Measurement

Rasch measurement can be used to measure any aspect of a situation. The greatest advantage of employing the Rasch model is how flexible and successful it has been in analyzing data over the years (Fox & Jones, 1998; Kim & Wilson, 2010; Liu, Minsky, Ling, & Kyllonen, 2009; MacMillan, 2000a; McHorney, Haley, & Ware, 1997). One thing that allows the Rasch model to be so effective is that the model controls the researcher's thinking (Linacre, 2010). Due to the expectation that the data must fit the model, large sample sizes are not required for Rasch analysis. Researchers are required to examine how well the data fits the model by examining the differences between the observed scores and the expected scores (Lochhead, 2009). McHorney, Haley, and Ware (1997) explain that when data are missing, researchers can use the expected score information to calculate the missing data. The robust nature of the Rasch model seems highly appropriate to application selection because university programs never know exactly how many applicants are going to apply to the program each year. The principles of Rasch measurement are advantageous in situations, like applicant selection, where it may not be feasible for all raters to evaluate every applicant on each item of measurement.

Fox and Jones (1998) explain that a Rasch analysis produces a set of Infit and Outfit values, also referred to as Fit statistics, for all facets of a data set. This means that for each applicant, item, and rater involved in this research study estimated parameters were calculated. These Fit statistics are useful to researchers as they provide information about which applicants, items, or raters behaved unexpectedly (Fox & Jones, 1998). Fit statistics have a mean-square value of 1.0 and have a positive infinite range (Linacre, 1995). Mean-square values that are greater than 1.0 are labelled "underfit", while mean-square values less than 1.0 are labelled "overfit" (Linacre, 2010). There is no agreed upon range for mean-square values; however, Wright and Linacre (1994) state that a reasonable range would be between 0.6 and 1.4, although it can vary depending on what exactly is being measured. Engelhard (1992) suggests that an acceptable range for Infit and Outfit statistics is 0.5 to 1.5. Linacre (1999) warns researchers to observe extreme caution with Fit statistics that are

greater than 2.0 because Fit statistics that are greater than 2.0 are so unexpected that there is hardly any useful information that can be reliably inferred. R. M. Smith (2004) as well as others are deliberately conscious of paying attention to the *t*-tests, $|t| \ge 2$ that accompany Infit and Outfit values as they feel the interpretation means more than a specific number generated by the computer software. Infit mean-square statistics provide information about the unexpected inliers and examine the region where a person's ability generally is. Outfit mean-square statistics provide information about the unexpected outliers and would be more sensitive to situations where a person answered a really easy item incorrectly or a really difficult item correctly (Linacre, 2010). If for some reason the data were problematic, Fit statistics would be the first place the researcher would detect the problem.

The Rasch model is more descriptive of the problems that may not be observed if the researcher was using another model. For instance, an ANOVA could be used to provide the same sort of analysis; however, an ANOVA would have a difficult time adjusting for incomplete data and raw scores that have not been standardized (Lunz, Wright, & Linacre, 1990). Linacre (1997) suggests that even incomplete data are not a problem for the Rasch model because usually a best estimate can be calculated. It appears that at the present time, the Rasch model is the optimum tool used for measurement in the human sciences, more specifically Education and Health Sciences (Bond & Fox, 2007).

Dichotomous Rasch Model. When Georg Rasch (1980) initially described the basis of his model in 1953, he used correct/incorrect, true/false, and yes/no examples to illustrate a person's responses to individual items. Rasch hypothesized that statistics generated from the testing process could be filtered down to person ability and item difficulty. He believed that the probability of getting a correct answer was equal to the ratio between person ability and

item difficulty. The dichotomous Rasch model is known for being the basic starting point for Rasch analysis. The dichotomous Rasch model for person ability and item difficulty is shown below:

$$\log_{e}(P_{ni}/1-P_{ni}) = B_{n} - D_{i}$$

where

 $log_e(P_{ni}/1-P_{ni}) = log-odds$ of person n succeeding on item i P_{ni} = the probability that person n correctly answers item i $1-P_{ni}$ = the probability that person n incorrectly answers item i B_n = ability of person n

$$D_1$$
 = difficulty of item *i* (Linacre, 2010)

Georg Rasch employed a natural log-odds ratio that represents numeric values using two symbols, 1 (indicating success) and 0 (indicating failure). He used this idea to create an additive system, rather than a multiplicative system that was used by mathematicians previously (Wright, 1997). The Rasch model transforms qualitatively ordered data into interval data using the mathematical principles of logarithms. Logit scores are the units produced by this conversion. When $P_m = 1 - P_{ni}$ (the chance of success on an item = the chance of failure on an item) the logit value will have a mean of 0.0. The logit scale is an interval scale of measurement where each individual logit unit has meaning because of the measurable differences that can be observed (Bond & Fox, 2007). This transformation from ordinal data to interval data is necessary because researchers interested in measurement need observable measures, not raw scores to make inferences about the data. Ordinal and interval data both reflect logical order. Hence, the main difference between ordinal and interval is that interval measures have equal units of measurement, and those equal units of

measurement imply equal differences in value (Hurlburt, 2006). The reason why inferences cannot be made from ordinal data is because raw scores include unwanted parameters (Wright, 1997). Through Rasch analysis these raw scores are calibrated so that the logit scores create a distribution of person abilities and item difficulties that can be compared and measured (Lochhead, 2009).

Rasch-Andrich Rating Scale Model. In 1978, David Andrich took the principles of the dichotomous Rasch model and created the rating scale model, which he believed was made possible through a series of Rasch dichotomies (Linacre, 2010). This extension of the dichotomous Rasch model was designed to deal with items that have more than two response options. A common application of this model is observed with the use of Likert scales (Strongly agree, Agree, Neutral, Disagree, Strongly disagree) where the distance between each response is designed to be the same (Linacre, 2010). The Rasch-Andrich rating scale model for person ability and item difficulty is shown below:

$$\log_{e}(P_{nij}/P_{ni(j-1)}) = B_{n} - D_{i} - F_{j}$$

where

P _{nij}	= probability of person n scoring at level j on scale i	
P _{ni(j-1)}	= probability of person n scoring at level j - l on scale i	
B _n	= ability level of person n	
D _i	= difficulty of level of item i	
Fj	= difficulty of the step from level j -1 to j	(Linacre, 2010)

Andrich (1996) explained that the scaled responses are similar to dichotomous measurement; however, the only difference is that the rating scale model partitions responses into intervals across the latent unidimensional linear construct. **Rasch-Masters Partial Credit Model.** By 1982, Geoff Masters advanced the rating scale beyond the work of Andrich to reflect partial credit for partial correctness, specific to each individual item (Linacre, 2010). According to Masters (1982), the rationale for creating partial scoring is to provide as much information as possible about a person's overall ability. Think of being asked to solve a math problem that requires five steps. Say a person calculated all the steps correctly with the exception of a slight calculation error in the final step causing them to get the final answer wrong. If the question was scored using a dichotomous model, the person would get the question wrong because they failed to get the final answer correct. With a partial credit model, the person would be able to get 4/5, suggesting that he or she was close to getting the final answer correct. In this version of the rating scale, each individual item has the freedom to vary in its number of estimates (Lochhead, 2009). The Rasch-Masters partial credit model for person ability and item difficulty is shown below:

$$\log_{e}(P_{nij}/P_{ni(j-1)}) = B_{n} - D_{i} - F_{ij}$$

where

P _{nij}	= probability of person n scoring at level j on scale	i
$P_{ni(j-1)}$	= probability of person n scoring at level j - l on scal	e i
B _n	= ability level of person n	
D1	= difficulty of level of item i	
F _{ıj}	= difficulty of the step from level j -1 to j in the rating scale equation spec	
	to the measure of difficulty for item <i>i</i>	(Masters, 1982)

This equation shows that each item has its own rating scale that is specific to the difficulty level of that particular item. If the researcher was looking solely at how the

applicants were performing on various items, then a partial credit model could be used. However, because the applicants were being rated by faculty and students on these various items the Many-Facet Rasch Model needs to be employed.

Many-Facet Rasch Model. Research over the last couple of decades has demonstrated that the Many-Facet Rasch Model (MFRM) can be successfully applied in various settings (Chang & Chan, 1995; Engelhard, 1992; Kim & Wilson, 2010; Linacre & Wright, 2004; MacMillan, 2000a; Smith & Kulikowich, 2004). The MFRM is an extension of the original Rasch measurement model as it goes beyond person ability and item difficulty to measure other factors that interact with a testing situation. The MFRM was developed from the work of John Michael Linacre (Bond & Fox, 2007). An example of what a three-facet Rasch model would look like is featured below:

$$Log_e(P_{nijk}/P_{nij(k-1)}) = B_n - D_i - C_j - F_k$$

where

 D_1

Pnuk = probability of examinee *n* being graded k by judge *j* on item *i* $P_{nu(k-1)}$ = probability of examinee *n* being graded *k*-1 by judge *j* on item *i*

$$B_n$$
 = performance measure of examinee *n*

$$D_1 = difficulty of item i$$

$$C_1$$
 = severity of judge *j*

 $\mathbf{F}_{\mathbf{k}}$ = difficulty of grading Step [category] k relative to Step [category] k-1

```
(Lunz, Wright, & Linacre, 1990)
```

Using the Many-Facet Rasch Model, Chang and Chan (1995) examined the functional independence measure of patients in a stroke rehabilitation program; while MacMillan (2000b) applied it towards assessing Curriculum Based Measurement reading scores.

Nevertheless, one of the places where the Many-Facet Rasch Model seems to be most successful is with studies that have observed raters and judges (Engelhard, 1994; Linacre, 1997; Lunz, 1999; Linacre, Wright, & Lunz, 1990; O'Neill, 1999). It is worthwhile to include rater behaviour in the assessment of a person's overall ability because the severity of individual raters could have a significant impact on how a person's ability is articulated.

The Effect of a Rater's Influence

Human beings all perceive the world subjectively through their own set of standards. This subjectivity is significant in the context of looking at rater behaviour because an individual's rating of another person's ability can be difficult to measure. Some studies like the one conducted by Liu and colleagues (2009) have put forward the unsubstantiated view that as long as raters share a similar background and are motivated by what they are doing, then rater variability is not an issue. Other studies carried out by Engelhard (1994) and O'Neill (1999) clearly demonstrate that each rater differs significantly based on his or her own personal standards of excellence. Research in this area has demonstrated that providing training to raters does not alter rater evaluations because each rater's level of severity seems to be engrained in his or her personal view of what it is being assessed (Lunz, Wright, & Linacre, 1990). However, some researchers state that training raters to become aware of the effects, biases, and errors involved in the process of rating minimizes rater errors in a variety of settings (Edward W. Wolfe, personal communication, April 27, 2010).

Leniency/Severity effect (or generosity error). The leniency is used to describe the behaviour of individuals who rate above the midpoint of the scale, while severity is used to describe the behaviour of those who rate below the midpoint of the scale (Myford & Wolfe, 2004a). Myford and Wolfe (2004a) describe these effects as tending to occur when raters

know, or have been able to identify in some way with, whom they are rating. There are a few ways that a researcher could detect rater effects from a Rasch analysis. Myford and Wolfe (2004b) suggest looking at the following statistical indicators for signs of rater leniency or severity effects: fixed chi-square values for raters, rater separation ratio and reliability, Wright variable map, rater severity, and rater fair average measures. Alternatively, Engelhard (1994) suggests looking at the Fit statistics for each rater; Fit statistics could be used to determine how much each rater would need to be calibrated.

Central tendency effect. Central tendency effect is when a rater avoids using the outermost categories (Linacre, 2010). It can also occur when a rater overuses the middle categories (Myford & Wolfe, 2004a). Central tendency effects often occur because the rater is afraid to make a mistake. The problem with raters that overuse the middle categories is that every applicant they rate looks average and thus, their ratings become information poor. Myford and Wolfe (2004b) suggest looking at the following statistical indicators for signs of central tendency effects: fixed chi-square values for applicants, applicant separation ratio and reliability, rater Fit statistics, rating scale category thresholds and probability curves for raters.

Restriction-of-range effect. Range restriction is closely related to central tendency in that it occurs when all of the raters avoid using certain categories or have scores that are clustered together, not necessarily around the midpoint (Myford & Wolfe, 2004a). One of the problems with restriction-of-range effects is that the rating scale is not fully represented, which means that the item to which an applicant is being rated is using a different rating scale than initially intended. Myford and Wolfe (2004b) suggest looking at the following statistical indicators for signs of restriction-of-range effects: the standard deviation across all

applicants for a specific item, ANOVA interactions, frequency counts for how many times each rater used each point on the scale, and probability curves.

Halo effect. Halo effect can occur when the evaluation given by the first rater influences the ratings given by the next rater (Linacre, 2010). Engelhard (1994) defines the halo effect as the rater assessing a person's ability holistically rather than on an item-by-item basis. Myford and Wolfe (2004a) suggest that out of all the rater effects, the halo has been most studied and received the largest amount of attention in the literature. Halo effects result when a rater is unable to separate independent aspects of a person's behaviour. Myford and Wolfe (2004b) suggest looking at the following statistical indicators for signs of restrictionof-range effects: fixed chi-square values for items, item separation ratio and reliability, rater Fit statistics, rater-by-item interaction analysis.

Myford and Wolfe (2004a) have identified other rater effects that are less likely to be encountered such as randomness, inaccuracy, logical error, contrast error, influences of rater biases, beliefs, attitudes, and personality characteristics, influences of rater/applicant background characteristics, proximity error, primacy error, and order effects. These effects are also less prevalent in the research literature probably because they are more difficult to measure.

With a better understanding of how the effects of a particular rater can bias the evaluation of individuals, researchers are better able to attain a more accurate account of a person's true abilities. Without some measure adjusting for rater variation, a person's ability score could be heavily influenced by whether the person rating them was severe or lenient (Lunz et al., 1990). It would be ideal if all raters on a committee could rate each individual separately in a full crossed design; however, in many cases that is not the most feasible approach

considering the quantity of applications certain programs receive and the fast turn around time that is generally required. According to Linacre (1997), fully crossed designs are not essential: "the only requirement on the judging plan is that there be enough linkage between all elements of all facets that all parameters can be estimated without indeterminacy with one frame of reference" (p. 1). Therefore, in situations where it is not feasible to have every rater evaluate all applicants, Rasch analysis would still be possible as long as there is some overlap among raters.

Generalizability Theory

Over the last decade there have been countless studies carried out that have assessed and analyzed a person's ability using Generalizability Theory (Atilgan, 2008; Oosterveld & ten Cate, 2004; Pedersen, Hagtvet, & Karterud, 2007; Winne et al., 2006). Atilgan (2008) used G-Theory to assess and score applicants who would later be selected as students for music education programs. Oosterveld and ten Cate (2004) used G-Theory to assess the applicant selection procedure for students wishing to attend medical school. G-Theory describes how reliable one person's score is when generalized to the greater universe of all scores. G-Studies are designed to provide researchers with information about the sources of variation that contribute to error in measurements by providing estimates for each source of error (Shavelson & Webb, 1991). By examining all of the identified main effects and interactions of facets that are involved in a study individually, researchers are able to account for the unexplained sources of variability, and thus produce a G coefficient that reflects the true amount of variance associated with a person's score (Shavelson & Webb, 1991). Every G-Theory analysis produces a set of G coefficients, one for relative decisions and one for absolute decisions.

Relative and absolute decisions. Relative and absolute decisions are made based upon how the researcher wishes to generalize his or her findings. Relative decisions involve interpreting an individual's overall placement in relation to other individuals (Kieffer, 1999; Shavelson & Webb, 1991). Within this particular study which looked at the application selection process of MEd-Counselling applicants, all the G-Study decisions would be relative given that the researcher was looking at how well applicants, raters, and items place in relation to each other. Absolute decisions are made when an individual's level of performance is determined by achieving a minimal level regardless of how it sits in relation to other individuals (Kieffer, 1999; Shavelson & Webb, 1991). A good example of an absolute decision within G-Theory analysis is how individuals are evaluated for their learner driver's license; once an individual gets 40 questions correct, they pass regardless of how many other people also got 40 questions correct on that day.

Random and fixed facets. A crucial concept associated with G-Theory is the distinction between random and fixed facets. A facet is considered random if the researcher is willing to interchange one for another (Kieffer, 1999). For example, say that a researcher interested in examining rater behaviour has selected two second year counselling students to rate applications from the population of all second year counselling students. If the researcher was willing to exchange those two students initially selected for two other students from the population of second year counselling students, then the rater facet would be considered random. Theoretically in this case, the rater behaviour of the students selected could be generalized to the population of which the students were drawn from. A facet is described as fixed when the researcher has included exactly those in the area of interest and is not willing to interchange them for another (Kieffer, 1999). Using the example mentioned previously, if the researcher was inflexible with the students who were selected as raters, then the rater facet would be considered fixed. Gender (male and female) is another example of a fixed facet as there are no other genders of interest available to the researcher.

Decision studies. Within the framework of G-Theory the researcher takes on the role of a decision maker. The first part of a G-Study is to estimate variance components that would allow the researcher to examine the differences between an observed score and the universe of all possible scores (Matt, 2010). The second part of a G-Study is to make decisions about the optimal measurement condition which would allow the researcher to sufficiently generalize the results obtained from the G-Study (Pedersen, Hagtvet, & Karterud, 2007). Decision studies provide indicators for how a study could be refined and better developed in the future. Within the context of this study, the decision studies would reflect the number of raters and items that would be required to make the results of the study reliable based on the variance components of the G-Study.

Rasch Meets Generalizability Theory

Linacre (1989) was one of the first researchers to put the Many-Facet Rasch Model (MFRM) together with Generalizability Theory (G-Theory). MacMillan (2000b) introduced the combination of Classical Test Theory, MFRM, and G-Theory. MacMillan found that all approaches were effective in detecting rater variability. However, the MFRM identified more variation than G-Theory did. G-Theory considers not only the facets, but also interactions among the facets whereas the MFRM assumes there are no interactions and treats each facet independently (MacMillan, 2000b). Countless researchers have been exploring the relationships between multiple methods to enhance the ability to fully answer research questions. The combination of the MFRM and G-Theory became popularized around 2004/2005. Using the MFRM and G-Theory combination, Smith and Kulikowich (2004) assessed the problem-solving skill level of fourth grade students. They found it useful to use both MFRM and G-theory with the same data set because although both measurement models provide information about variability that exists among facets, the approach used to obtain such information differed greatly (Smith & Kulikowich, 2004). Linacre and Wright (2004) published a book chapter that explored the construction of the MFRM and G-Theory by using data that required judges to rate examinees on various items. Linacre and Wright (2004) concluded:

Generalizability theory (G-Theory) and Many-facet Rasch measurement (MFRM) appear to be competing methodologies aimed at solving the same empirical problems. But this is not the case. Though the data sets specified by the two methodologies may be similar, or even identical, their purposes are fundamentally different. (p. 312)

Furthermore, G-Theory attempts to correct for variability in future sets of data collection, while the MFRM attempts to correct for variability in the current set of data (Linacre & Wright, 2004). Given the nature of the present study, the Rasch model seems to fulfill the needs of the institution better than G-Theory at this time. Sudweeks, Reeve, and Bradshaw (2005) suggest that G-Theory is a holistic approach to analyzing data as it examines main effects and interactions, while the MFRM is a narrow approach that focuses on the individual facets for analysis. Both G-Theory and the MFRM have strengths when analyzing data, and Sudweeks and colleagues (2005) propose that using both forms together yields a more comprehensive analysis of what is really occurring within a given data set. Most recently, Kim and Wilson (2010) used data from individuals assessing student writing ability to expand on the notion that the MFRM and G-Theory are more than alternative methodologies to measuring data. Although both measurement methodologies will provide

information about the applicants, items, and raters, Kim and Wilson (2010) suggest that the researcher needs to clearly outline what exactly it is that they are interested in studying. If the researcher is interested in looking at how groups are behaving, then G-Theory would be most helpful; however, if the researcher is concerned with individual performance, then the MFRM or another item response model should be employed.

Precise and accurate measurement requires the right set of tools. Many measurement models, like the Rasch model, have evolved to meet the needs and demands of the facets that individuals want to measure. Although the expansion of research and its associated literature has grown to the point where researchers can measure almost anything, error and bias are still likely to occur. The more complex the object that researchers are set out to measure, the more creative researchers need to be in their measurement approach. Combining different measurement methodologies is one way in which researchers can gain a more holistic approach and capture the true representation of what they wish to measure.

Chapter Three: Research Methods

The Many-Facet Rasch Model (MFRM) and Generalizability Theory (G-Theory) are two highly researched and well established measurement methodologies that will be used to assess the MEd-Counselling application selection process. These two methodologies were particularly applicable considering the intended purpose of this study. To begin the analysis, it was interesting to look at how well the items used as pre-admissions criteria were measuring an applicant's overall level of ability. Another worthwhile area to examine was the rating behaviour of the members on the application selection committee from both individual and group perspectives. The information generated from the evaluation of specific raters would determine how much each rater influences a particular applicant's chances of being offered a letter of acceptance. Research has found that most institutions are willing to accept that there are differences in scoring from one rater to another and they may even try to correct for this by selecting raters who share a similar background (Liu, Minsky, Ling, & Kyllonen, 2009). However, studies examining interrater reliability argue that regardless of any attempts made, there still exist significant differences between raters (Lunz, Wright, & Linacre, 1990). The present study tried to examine exactly where these differences were located. For instance, looking at whether rater differences were exhibited among genders (male and female), education level (faculty and students) or discipline (counsellors and non-counsellors) could have important implications for determining which individuals should sit on the selection committee in the future. Finally, using the MFRM and G-Theory to measure these differences alleviated any concerns about the overall effectiveness of the application selection process and ensured fairness for all applicants applying to the program.

Participants

Raters. The participants for this study were three faculty members and two graduate students. All of the participants agreed to take part in this study when they signed a consent form to participate as a rater (to view the consent form please see Appendix B). The faculty sample consisted of three faculty members from the School of Education. All three faculty members teach in the counselling program, although only two out of the three faculty members were educated as counsellors. The graduate students consisted of two graduate students currently in the MEd-Counselling Program. The two graduate students were both near completion of their degrees and are expected to have graduated before the successful applicants enter the program in September 2010. The researcher is one of the two students; the other student was approached after agreement between the supervisor and the researcher.

In order to ensure anonymity of applicants for the purposes of this research, all application packages were stripped and coded for the participants by the Chair of the selection committee. The five participants agreed to examine the data from individuals who applied to the MEd-Counselling program for intake in September 2010. Following the UNBC's established procedure, all applications were collected by the University's registrar's office. Applicant packages of individuals who met the GPA requirement (as well as those individuals who did not meet the GPA requirement, but were specifically requested by the Counselling Coordinator) were then forwarded to the Counselling Coordinator in the School of Education who checked them over and prepared them for the selection committee.

Applicant pool. The probable applicant pool consisted of applicants from two campuses: Prince George and the Northwest region. The population applying was roughly 80% female and 20% male. Applicants ranged in age from 22 years old to 55 years old. Most applicants
obtained a Bachelor's degree in Psychology, Social work, Criminology, or Education. Finally, the level of relevant work experience of those applying to the program ranged from volunteer experience in the helping arena to those who have been employed in the counselling field for over 30 years.

Issues of access to applicant pool. There were no issues of access to the applicant data that the participants rated as consent was inherent with the application (implied consent). Individuals who applied to the University have given consent for the information they supply to be used for research purposes when they apply to the program ("UNBC Graduate Calendar," 2009).

Ethical Considerations

Ethical issues are always a critical consideration when conducting research. The challenge of ethics with this research was ensuring that the principle of confidentiality was honoured for the protection of the applicant data set. To ensure that confidentiality was maintained throughout this research study, identification numbers rather than names were used to code all of the applicant packages. The Chair of the committee stripped and coded the data so that all raters were given application packages with no identifiers. The average number of applications that the Counselling Program in the School of Education had typically received in past years was roughly 30-40, suggesting that there should be no added risk of successful applicants being later identified by the student raters.

Both of the student raters used in this study have successfully completed the counselling ethics course at UNBC and are fully expected to uphold the values of confidentiality set out by the licensing organization, Canadian Counselling and Psychotherapy Association; this includes maintaining confidentiality in both therapeutic and research settings. Furthermore, both of the students who rated applications have completed all course requirements and are expected to graduate before September 2010; therefore, neither of the student raters would be expected to encounter any of the successful applicants as a student in the future.

Instruments

An application package consisted of (a) a grade-point-average (GPA), (b) relevant degree information (c) written evidence of involvement with people in appropriate settings, (d) a written personal statement, and (e) three letters of reference. This pre-admissions criterion was developed by faculty members from UNBC's School of Education. Corey, Corey, and Callanan (2003) recommended all these pre-admissions criteria as suitable for screening potential students for a counselling program. For the present study, the application packages contained all of the same information that was collected in the previous years to select counselling students. Members of the Counselling Program application selection committee rated all of the information, with the exception of GPA, on a series of 5-point scales. Every member of the selection committee, as well as both of the student raters, rated and scored the entire application package for each applicant that applied to study at the Prince George campus. An overall score based on all of the application criteria was used to rank the applicants. The rank ordering information generated throughout this study was given to the faculty members to use for their final decisions about who would be offered a seat in the program. Although the two student raters were participants in the research and rated the applicants for the purpose of this study, the final decisions as to letters of offer came from the admissions committee that did not include the student raters.

Procedures

This study was reviewed and supported by the UNBC Research Ethics Board. The Research Ethics Board stated that ethics approval was not required as they found this study to be a typical program evaluation that would not interfere with the established protocol for selecting MEd-Counselling applicants (See Appendix C). Upon receiving this decision both verbally and in writing, the researcher contacted the Chair of the application selection committee who held all of the applications that were received. The Chair of the application selection committee had these applications photocopied with all identifiers removed for each of the committee members. All members of the selection committee, as well as the two student raters who participated in this research study, were given a brief training session where the established selection process was explained. The researcher and her supervisor described the 5-point scales used for rating the applicants (See Appendix D). The researcher also briefly explained to the participants some of the most common rating behaviours that have been shown to be problematic.

Each rater participant in this study was given copies of all the application packages that needed to be evaluated. The students' application packages were exactly the same as the ones given to the faculty selection committee members. All applications were read and scored within a two-week period as agreed upon by the selection committee. Once all five raters finished scoring every application package, the packages were returned to the Chair of the selection committee. Each applicant GPA and all rater scorings were entered into an EXCEL file by the researcher.

Measures used for Analyzing Data

There are different statistical programs that could be used to analyze the data, such as Microsoft EXCEL, SPSS, or FACETS. For the purpose of this study, the data obtained from the raters was compiled using EXCEL and analyzed using FACETS (Linacre, 1996) and EDUG (Swiss Society for Research in Education Working Group, 2010).

Rasch analysis. The research design was a fully crossed three-facet Rasch analysis examining person ability by rating the quality of their counselling application, the difficulty of the items on the 5-point scale, and the severity of all the raters on the selection committee. The many-faceted Rasch rating scale model used for this analysis of the data set was the same one as described earlier in the literature review. This measurement model was used for the analysis because it allowed the researcher to examine interactions between multiple facets. Each facet was examined to see the level of influence it had on the probability of a particular applicant scoring the way they did on specific items by various raters.

Generalizability analysis. In addition to the Rasch analysis, the research data was examined using a fully crossed two-facet Generalizability analysis. For the purpose of this study, item difficulties and rater behaviour are the two facets that were analyzed. The applicants are described as the objects of measurement in G-Theory; therefore, a three-facet Rasch study and two-facet G-Theory study are for all intents and purposes the same analysis. A two-facet G-Theory design is associated with six sources of variability that can be examined as follows:

Source of Variability	Type of Variation
Persons (p)	Universe-score variance (object of measurement)
Raters (r)	Constant effect for all persons due to stringency of raters
Items (i)	Constant effect for all persons due to their behavioural
	inconsistencies from one item to another
<i>p</i> * <i>r</i>	Inconsistencies of raters' evaluation of particular persons'
	behaviour
p * i	Inconsistencies from one item to another in particular persons'
	behaviour
r * i	Constant effect for all persons due to differences in raters
	stringency from one item to another
p * r * i, e	Residual consisting of the unique combination of p, r, o;
	unmeasured facets that affect the measurement; and/or random
	events

(Shavelson & Webb, 1991, p. 9)

G-Theory allowed the researcher to partition different sources of variability that exist within measurement situations. This process was mathematically possible using the same logic and mechanics of a factorial ANOVA (Shavelson & Webb, 1991).

Chapter Four: Results

Many-Facet Rasch Analysis

The results of the many-facet Rasch analysis are shown in Figure 1. The far left column titled "Measr" is the logit scale used to measure all of the facets within the design. The second column is the distribution of the applicants; most of the applicants were situated within the 0 to 2 region on the logit scale, indicating that they were proficient applicants. The third column contains the program status: full-time or part-time studies. The fourth column is the rater facet. Notice that all of the raters were positioned around the 0 logit mark. Those raters above 0 logits would be considered more severe, while those below 0 logits would be considered less severe. The raters will be examined individually later in this section. The fifth column represents the item difficulties; more difficult items are in the positive logit region and the less difficult items are in the negative logit region. The item difficulties will also be discussed later in this section. The final column "S.1" shows the ratings on the 5-point scale.

	Measr	+Student	-Program	-Rater	-Items	S.1
+	- 3 -	+ +		+ -	+	+(5) +
		-t-				
		*				
					1	
Ì					1	
İ					l	
İ		**			ĺ	i i
1						
+	2 +	+ -		+ -	∔ :	+ +
		*				
		*				
		* * * *				
		* * *			ļ	
İ		* * * *		ł	İ	İ İ
		* * * * *				
		* * * * *				4
	1	****		1		
+	· ·	- ^ ^ ~ · · · · · · · · · · · · · · · · ·		+ - 	+ 	+ +
		* * * *			4	
i						
Ì		* * * *		ĺ	ĺ	İ
ļ		*				
		* *			8	
		*		R1 R3	6	
*	0 *	k * 1	PGFT PGPT	* R2 *	*	· · ·
				R4		
	ļ			R5	3	3
		*				
					9	
	l				 7	
					,	
						i i
İ	İ				5	İİ
+	-1 +	- 4	- 4	+ -	+ ·	+(1) +

Figure 1. Wright variable map for relationships among facets for Prince George applicants.

The many-facet Rasch analysis was completed using FACETS software, version 3.03 (Linacre, 1996). According to Linacre (2010), the data need to fit the Rasch model if they are to support linearity. A unidimensional Rasch analysis operates under the assumption that multiple observations can be viewed as one theoretical construct (Bond & Fox, 2007). By conducting a many-facet Rasch analysis, the researcher was able to see that all of the criteria items used to assess prospective MEd-Counselling students (degree, writing ability, goals, work experience, referee quality and suitability) fit within a unidimensional construct. A summary of the item characteristics and their facet statistics are located in Table 1.

Table 1

Obsvđ	Fair		Model	Infit	Outfit			
Average	Average	Measure	S.E.	MnSq ZStd	MnSq ZStd	Nu Items		
3.8	2.96	.26	.08	0.6 -4	0.6 -4	1 Degree		
3.8	3.02	.19	.08	1.0 0	1.0 0	2 Writing Ability		
4.1	3.38	24	.09	0.9 0	0.9 0	3 Fit of Goals		
3.4	2.40	.84	.07	1.5 4	1.5 4	4 Work Experience		
4.4	3.82	87	.10	1.2 1	1.1 1	5 R1:Suitability		
3.9	3.08	.13	.08	0.9 0	1.0 0	6 R1:Quality		
4.3	3.66	63	.09	1.2 2	1.2 2	7 R2:Suitability		
3.7	2.84	.38	.08	0.8 -2	0.8 -2	8 R2:Quality		
4.2	3.46	35	.09	1.2 1	1.2 1	9 R3:Suitability		
3.8	2.93	.30	.08	0.8 -2	0.8 -2	10 R3:Quality		
Adj S.D48 Separation 5.67 Reliability .97								
Fixed (al	Fixed (all same) chi-square: 320.7 d.f.: 9 significance: .00							
Random (r	Random (normal) chi-square: 9.0 d.f.: 8 significance: .34							

Items	Measurement	Report
-------	-------------	--------

Recall from the literature review that Engelhard (1992) suggests that an acceptable range for Infit and Outfit statistics is 0.5 to 1.5. R. M. Smith (2004) as well as others are deliberately conscious of paying attention to the *t*- tests, $|t| \ge 2$. High Infit and Outfit statistics may be viewed as an indication of multidimensionality. The work experience item had the highest Infit and Outfit value of 1.50 with a t = 4.0; this value borders on what is considered an acceptable range for Infit and Outfit statistics. Typically, one misfitting item would not normally be considered an indication of a lack of unidimensionality or model fit. However, on-going discussion among the primary researcher and the other raters indicated differing views of the work experience and writing ability items. All raters were aware that there were two types of applicants, those applying for full-time studies and those applying for part-time studies. For this particular reason, the applicant pool was divided according to full-time or part-time status and the item analysis repeated.

Applicant pool. By dividing the applicants into two groups, those wishing to pursue fulltime studies and those wishing to pursue part-time studies, it became clear that there are two populations existing within the applicant sample. The item summary for the full-time students is featured in Table 2 and the item summary for the part-time students in Table 3.

Table 2

Obsvđ	Fair		Model	Infit	Outfit		
Average	Average	Measure	S.E.	MnSq ZStd	MnSq ZStd	Nu	Items
3.9	3.12	.08	.10	0.5 -5	0.5 -5	1	Degree
3.9	3.15	.04	.10	1.0 0	1.1 0	2	Writing Ability
4.1	3.35	24	.10	1.0 0	1.0 0	3	Fit of Goals
3.0	1.96	1.42	.08	1.2 1	1.2 1	4	Work Experience
4.4	3.76	89	.12	1.2 1	1.2 1	5	R1:Suitability
3.8	3.06	.16	.10	1.0 0	1.0 0	6	R1:Quality
4.3	3.68	76	.11	1.2 1	1.2 1	7	R2:Suitability
3.7	2.81	.46	.09	0.9 0	0.9 -1	8	R2:Quality
4.2	3.59	61	.11	1.2 2	1.2 1	9	R3:Suitability
3.7	2.91	.34	.10	0.9 -1	0.9 -1	10	R3:Quality
Adj S.D.	.64 Sej	paration	6.33 Re	Liability .9	8		
Fixed (a)	ll same) ci	hi-square:	454.5	1.f.: 9 sign	ificance: .00		
Random ()	normal) ch	i-smare:	9.0 đ.f	. 8 simifi	cance: .34		

Full-time Applicants Items Measurement Report

The misfit for the work experience item now disappeared in both of the two analyses.

However, a new feature is apparent. The full-time applicants found the work experience

item to be the most difficult item (logit score = 1.42); they were rated lowest on this item while the part-time applicants found the work experience item to be the least difficult (logit score = -1.38) and were rated highest on this item. An item's level of difficulty, for the purpose of this study, was defined by the logit measure, which indicates the difficulty of endorsement by the raters on each particular item. The "Obsvd Average", shown in the far left column, gives the average of the raw observed scores. The second column titled "Fair Average" is the interval based adjustment of the observed average score as calculated from the linear transformation of the raw score.

Table 3

AverageAverageMeasureS.E.MnSq ZStdMnSq ZStdNu Items3.52.411.00.171.001.001Degree3.72.56.80.171.001.002Writing Ability4.23.2827.190.900.903Fit of Goals4.63.91-1.38.241.101.104Work Experience4.53.71-1.02.221.211.205R1:Suitability4.03.01.14.181.211.216R1:Quality4.23.3235.191.311.317R2:Suitability3.92.86.37.180.7-20.6-28R2:Quality3.92.88.34.180.8-10.8-110R3:Quality	Obsvđ	Fair		Model	Inf	it	Out	fit		
3.5 2.41 1.00 .17 1.0 0 1.0 0 1 Degree 3.7 2.56 .80 .17 1.0 0 1.0 0 2 Writing Ability 4.2 3.28 27 .19 0.9 0 0.9 0 3 Fit of Goals 4.6 3.91 -1.38 .24 1.1 0 1.1 0 4 Work Experience 4.5 3.71 -1.02 .22 1.2 1 1.2 0 5 R1:Suitability 4.0 3.01 .14 .18 1.2 1 1.2 1 6 R1:Quality 4.2 3.32 35 .19 1.3 1 1.3 1 7 R2:Suitability 3.9 2.86 .37 .18 0.7 -2 0.6 -2 8 R2:Quality 3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8	Average	Average	Measure	S.E.	MnSq	ZStđ	MnSq	ZStđ	Nu	Items
3.7 2.56 .80 .17 1.0 0 1.0 0 2 Writing Ability 4.2 3.28 27 .19 0.9 0 0.9 0 3 Fit of Goals 4.6 3.91 -1.38 .24 1.1 0 1.1 0 4 Work Experience 4.5 3.71 -1.02 .22 1.2 1 1.2 0 5 R1:Suitability 4.0 3.01 .14 .18 1.2 1 1.2 1 6 R1:Quality 4.2 3.32 35 .19 1.3 1 1.3 1 7 R2:Suitability 3.9 2.86 .37 .18 0.7 -2 0.6 -2 8 R2:Quality 3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8 -1 10 R3:Quality	3.5	2.41	1.00	.17	1.0	0	1.0	0	1	Degree
4.2 3.28 27 .19 0.9 0 0.9 0 3 Fit of Goals 4.6 3.91 -1.38 .24 1.1 0 1.1 0 4 Work Experience 4.5 3.71 -1.02 .22 1.2 1 1.2 0 5 R1:Suitability 4.0 3.01 .14 .18 1.2 1 1.2 1 6 R1:Quality 4.2 3.32 35 .19 1.3 1 1.3 1 7 R2:Suitability 3.9 2.86 .37 .18 0.7 -2 0.6 -2 8 R2:Quality 3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8 -1 10 R3:Quality	3.7	2.56	.80	.17	1.0	0	1.0	0	2	Writing Ability
4.6 3.91 -1.38 .24 1.1 0 1.1 0 4 Work Experience 4.5 3.71 -1.02 .22 1.2 1 1.2 0 5 R1:Suitability 4.0 3.01 .14 .18 1.2 1 1.2 1 6 R1:Quality 4.2 3.32 35 .19 1.3 1 1.3 1 7 R2:Suitability 3.9 2.86 .37 .18 0.7 -2 0.6 -2 8 R2:Quality 3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8 -1 10 R3:Quality	4.2	3.28	27	.19	0.9	0	0.9	0	3	Fit of Goals
4.5 3.71 -1.02 .22 1.2 1 1.2 0 5 R1:Suitability 4.0 3.01 .14 .18 1.2 1 1.2 1 6 R1:Quality 4.2 3.32 35 .19 1.3 1 1.3 1 7 R2:Suitability 3.9 2.86 .37 .18 0.7 -2 0.6 -2 8 R2:Quality 3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8 -1 10 R3:Quality	4.6	3.91	-1.38	.24	1.1	0	1.1	0	4	Work Experience
4.0 3.01 .14 .18 1.2 1 1.2 1 6 R1:Quality 4.2 3.32 35 .19 1.3 1 1.3 1 7 R2:Suitability 3.9 2.86 .37 .18 0.7 -2 0.6 -2 8 R2:Quality 3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8 -1 10 R3:Quality	4.5	3.71	-1.02	.22	1.2	1	1.2	0	5	R1:Suitability
4.2 3.32 35 .19 1.3 1 1.3 1 7 R2:Suitability 3.9 2.86 .37 .18 0.7 -2 0.6 -2 8 R2:Quality 3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8 -1 10 R3:Quality	4.0	3.01	.14	.18	1.2	1	1.2	1	6	R1:Quality
3.9 2.86 .37 .18 0.7 -2 0.6 -2 8 R2:Quality 3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8 -1 10 R3:Quality	4.2	3.32	35	.19	1.3	1	1.3	1	7	R2:Suitability
3.9 2.86 .37 .18 1.1 0 1.1 0 9 R3:Suitability 3.9 2.88 .34 .18 0.8 -1 10 R3:Quality	3.9	2.86	.37	.18	0.7	-2	0.6	-2	8	R2:Quality
3.9 2.88 .34 .18 0.8 -1 0.8 -1 10 R3:Quality	3.9	2.86	.37	.18	1.1	0	1.1	0	9	R3:Suitability
	3.9	2.88	.34	.18	0.8	-1	0.8	-1	10	R3:Quality
Fixed (all same) chi-square: 124.5 d.f.: 9 significance: .00	Random (r	normal) ch	i-square:	8.9 d.f	.:8 в	ignifi	cance:	.35		

Part-time Applicants Items Measurement Report

Among the part-time applicants, meeting the degree requirements appeared to be the most difficult item (logit score = 1.00) followed closely by level of writing ability (logit score = 0.80). In contrast, degree requirements and writing ability were of average difficulty (0.08 and 0.04 logits) for the full-time applicants. This seeming lack of invariance of item difficulty was judged to be legitimate population difference rather than a lack of fit of the

data to the Rasch model. Examination of the fit statistics separately for applicants applying to full-time or part-time studies (Table 5) yielded the same number of applicants who had Infit and Outfit mean squares that did not fall within what would be considered the acceptable range for fit statistics. When applicants were viewed as two distinct populations, both of the outliers, applicant 17 (Infit = 1.70, t = 3; Outfit = 1.70; t = 2) and applicant 27 (Infit = 2.10, t = 4; Outfit = 1.90, t = 3), were from the population applying for full-time studies and all of the applicants applying to part-time studies fit within the 0.50 to 1.50 range. However, when all of the applicants were considered as one population there were still two outliers, applicant 27 (Infit = 1.80, t = 3; Outfit = 1.70, t = 2) who applied for fulltime studies and applicant 48 (Infit = 1.80, t = 3; Outfit = 1.70, t = 2) who applied for parttime studies. Although examining the full-time and part-time applicants separately revealed some interesting results, there was no strong evidence to suggest that the two groups must be analyzed separately. The data demonstrate sufficient fit to the specified three-facet Rasch rating scale model. Therefore, the results featured below were generated from the 49 MEd-Counselling applicants competing for a seat at the Prince George campus.

Items. The Mean Square Fit indices have been discussed in relation to unidimensionality. The items are now described in more detail. The items, "R1, R2, R3, Suitability," are ratings of the suitability of the referees who provided a reference for the various applicants. Most referees were rated as well suited to comment on the appropriateness of the applicants. Conversely, the raters interpreted the referees' comments relatively severely, producing Fair Average measures of 2.84 to 3.08. Overall, the "appropriateness of the first degree," the "writing ability," and the "fit of the applicant's stated goals" with the nature of the counselling program were all items of average difficulty. The fixed (all same) chi-square of 320.7, df = 9 was found to be statistically significant (p < .005). This information solely suggests that the items are uniquely different in their level of difficulty from one another. Furthermore, all of the item scores together produced a separation ratio of 5.67 and a reliability coefficient of 0.97. The separation ratio produces a measure of the test items with the test error calculated in the measure (Fisher, 1992) and the reliability coefficient produces a measure of consistency for the item differences. Since the separation ratio and reliability coefficient are both high, these results indicate that each of the ten items used to evaluate applicants vary in their level of difficulty, thus capturing a wide range of applicant suitability to the MEd-Counselling program.

Raters. The rater measurement report shown in Table 4 describes the behaviour of each of the four raters. All of the Infit and Outfit statistics for the raters fell within an acceptable range. Myford and Wolfe (2004) suggest that in situations that involve high-stakes decision making, the fit mean square indices should be more stringent, adjusted to 0.8 to 1.2 in this case for judges. The Infit scores for the raters ranged from 0.80 to 1.30 and the Outfit scores ranged from 0.80 to 1.20. However, one student rater and one faculty rater both displayed ratings (Infit = 0.8; Outfit = 0.8) that would be "cramped" or "information poor" by the t-test criterion: -3 and -2 respectively for Infit scores and -2 and -3 respectively for Outfit scores. The other student rater demonstrated opposite rating behaviour (Infit score = 1.30, t = 4 and Outfit score = 1.20, t = 3), suggesting that the ratings given by this rater would be more erratic.

Obsvd	Fair		Model	Ini	Eit	Outfi	lt		
Average	Average	Measure	S.E.	MnSq	zStd	MnSq	ZStd	Nu	Rater
3.8	3.06	.15	.06	1.0	0	1.1	1	1	Faculty Counsellor (New)
3.9	3.18	.01	.06	0.8	-3	0.8	-2	2	Faculty Counsellor
3.9	3.11	.09	.06	1.0	0	1.0	0	3	Faculty Non-Counsellor
4.0	3.26	09	.06	0.8	-2	0.8	-3	4	Student Counsellor I
4.0	3.32	16	.06	1.3	4	1.2	3	5	Student Counsellor II
Adj S.D.	.10 Sej	paration	1.63 Re	liabili	ity .7	3			
Fixed (all same) chi-square: 18.1 d.f.: 4 significance: .00									
Random (1	Random (normal) chi-square: 4.0 d.f.: 3 significance: .26								

Rater Measurement Report

The most severe rater (R1) had a measure of 0.15 and the most lenient rater (R5) had a measure of -0.16, a spread of ± 0.16 logit off the mean with 2 S.E. being 0.12. This disparity suggests that the raters were fairly homogeneous when it came to rating the applicants. However, the fixed (all same) chi-square of 18.1, df = 4, was statistically significant (p < .005) which indicated rater differences. The separation ratio of 1.63 and the reliability coefficient of 0.73 indicate that the raters were somewhat reliably different. It does not matter for this study because the researcher used a fully crossed design, but in situations where the design is not fully crossed, it would be ideal if the reliability coefficient was lower. The lower the reliability coefficient, the more confident the researcher can be in the results, as a reliability coefficient of zero indicates that there is no difference between any of the raters (Sudweeks et al., 2005).

Applicants. The data used for this study consisted of raw scores from 49 MEd-Counselling applicants across ten items (rated on a scale of 1-5). The data from the applicants showed the results in Table 5.

Applicant Measurement Report

Obsvđ	Fair		Model	Inf	it	Out	fit		
Average	Average	Measure	S.E.	MnSq	ZStd	MnSq	ZStđ	Nu	Applicant
4.6	4.66	2.69	.26	0.7	-1	0.7	-1	1	11001
4.0	3.99	1.16	.19	0.9	0	0.9	0	2	11002
4.2	4.26	1.64	.20	1.5	2	1.4	1	3	11003
4.2	4.26	1.64	.20	1.3	1	1.2	0	4	11004
3.9	3.89	.99	.18	1.3	1	1.3	1	5	11005
4.0	3.99	1.16	.19	0.9	0	0.9	0	6	11006
4.1	4.09	1.33	.19	1.0	0	1.0	0	7	11007
4.1	4.15	1.45	.20	0.9	0	0.9	0	8	11008
3 6	3 63	59	17	0 7	-1	0.8	-1	9	11009
3.8	3.77	80	.18	1.0	0	1.0	0	10	11010
4 5	4 50	2 20	23	0.7	-1	0.7	_1	11	11011
4.0	4.50	1 48	20	0.9	Ô	n 9	Ō	12	11012
38	3 77	80	18	0.5	-1	0.5	0	13	11012
1.0	1 26	1 64	20	1 2	- 1	1 3	1	11	11010
30	4.20	1 09	.20	1.2		1.5		15	11014
2.9	2 01	1.09	.10	1 0	-2	1 1	-2	10	11015
2.9	3.91 2.16	1.02	.10	1.0	2	1.1	2	17	11017
3.1 4 1	3.10	03	.15	1.5	2	1.4	2	10	11017
4.1	4.09	1.33	.19	1.0	0	1.0	0	18	11018
3.8	3.//	.80	.18	0.9	0	1.0	0	19	11019
4.1	4.13	1.41	.19	0.6	-2	0.6	-2	20	11020
3.6	3.62	.56	.17	1.0	0	1.0	0	21	11021
4.1	4.15	1.45	.20	0.8	0	0.9	0	22	11022
2.9	2.89	33	.15	1.0	0	1.1	0	23	11023
3.7	3.67	.65	.17	0.8	0	0.9	0	24	11024
4.1	4.07	1.30	.19	0.5	-2	0.6	-2	25	11025
3.3	3.28	.11	.16	0.7	-1	0.7	-1	26	11026
4.2	4.26	1.64	.20	1.8	3	1.7	2	27	11027
3.8	3.77	.80	.18	1.5	1	1.4	1	28	11028
4.0	4.05	1.26	.19	1.1	0	1.2	0	29	11029
3.6	3.63	.59	.17	0.8	-1	0.8	0	30	11030
3.5	3.52	.42	.16	0.9	0	0.9	0	31	11031
3.8	3.85	.92	.18	0.6	-2	0.6	-2	32	11032
3.6	3.60	.53	.17	1.4	1	1.4	1	33	11033
4.3	4.34	1.82	.21	0.6	-2	0.6	-2	34	11034
4.3	4.30	1.73	.21	0.8	-1	0.8	-1	35	11035
3.5	3.52	.42	.16	0.5	-2	0.6	-2	36	11036
3.9	3.87	.96	.18	0.8	-1	0.8	-1	37	11037
3.9	3.95	1.09	.18	1.2	0	1.1	0	38	21038
4.5	4.51	2.24	.23	1.2	0	1.2	0	39	21039
4.0	4.01	1.19	.19	1.2	0	1.1	0	40	21040
4.2	4.19	1.52	.20	0.7	-1	0.7	-1	41	21041
4.1	4.09	1.33	.19	1.5	1	1.4	1	42	21042
4.0	3.99	1.16	.18	1.3	1	1.3	1	43	21043
3.9	3.91	1.02	.18	0.7	-1	0.7	-1	44	21044
4.1	4.15	1.44	.19	1.1	0	1.1	0	45	21045
4.1	4.17	1.48	.20	1.3	1	1.2	1	46	21046
4.0	3.99	1.16	.18	0.8	-1	0.8	0	47	21047
3.9	3.95	1.09	.18	1.8	3	1.7	2	48	21048
3.9	3,93	1.05	.18	1.4	1	1.4	-	49	21049
							<u> </u>		
Adj S.D.	.52 Sepa	ration 2.79) Reliabi	lity .	89				
Fixed (all	same) chi	-square: 436	5.6 d.f.:	48 si	gnifica	nce: .0	0		
Random (normal) chi-square: 47.5 d.f.: 47 significance: .45									

The Infit scores for the applicants ranged from 0.50 to 1.80, while the Outfit scores ranged from 0.60 to 1.70. As previously discussed, only two out of the 49 applicants had Infit and Outfit scores that did not fall within what would be considered the acceptable range for fit statistics. The applicant's ability measures ranged from -0.33 to 2.69 logits, mean = 1.14, SD = 0.56. The fixed (all same) chi-square of 436.6, df = 48, is statistically significant (p < .005). The separation ratio of 2.79 and the reliability coefficient of 0.89 indicate that the applicants are moderately heterogeneous, but nevertheless separable. As discussed earlier, these results are likely due to the combination of two groups: the full-time applicants and the part-time applicants.

Generalizability Analysis

The Generalizability analysis was completed using EDUG software, version 6.0 (Swiss Society for Research in Education Working Group, 2010). The design consisted of a two factor fully crossed design involving applicants (the object of measurement), raters (four individuals), and items (ten separate items). According to the Generalizability analysis, this study is considered to be a two-facet fully crossed design, with the two facets being the items and the raters. The applicants are not defined as a separate facet because they are considered to be the "object of measurement." With respect to the Rasch analysis, this study would be considered to be distinct facets. Using some of the same principles of traditional ANOVA, G-Theory uses variance components to represent the amount of error that comes from generalizing from a facets score to a universal score (Swiss Society for Research in Education Working Group, 2010). These variance components are shown in Table 6.

	Components								
Source	Variance Component	%	df	SE					
Persons (P)	0.095766	11.3	48	0.022414					
Raters (R)	0.004273	0.5	4	0.003557					
items (I)	0.066511	7.9	9	0.037895					
P*R	0.077503	9.2	192	0.007936					
P*I	0.257369	30.4	432	0.021688					
R*I	0.041624	4.9	36	0.010971					
P*R*I	0.303342	35.8	1728	0.010314					
Coef_G relative	0.86								
Coef_G absolute	0.85								

Estimated G-Study Variance Components

Variance components for items. The variance component for items reflects differences between each of the ten items. The variance component for the main effect of items was 0.067, accounting for approximately 8% of the total variance. Similarly, the variance component for the main effect of applicants was 0.096, accounting for approximately 11% of the total variance. Ideally, the variance component for applicants should be higher, indicating a more heterogeneous population. The more heterogeneous a population, the higher the value of the G coefficient. The reasoning for this correspondence is that in homogenous populations, raters have a harder time differentiating between applicants and thus, typically produce lower G coefficient values. Nevertheless, the variance component for the applicants were behaving differently from one item to the next.

Variance components for raters. The variance component for raters was 0.0043 (0.5%), indicating little variation due to rater differences. The rater-by-item interaction was also a comparatively small percentage 4.9% (0.042) of the overall variance, indicating that the raters used the scale consistently on each item. Interestingly, the variance component for the

raters-by-applicants interaction was 0.078 (9.2%), suggesting that the raters used the same standards, but disagreed on how they employed the standards for each of the applicants. These results were further reinforced through written comments and personal communication between the primary researcher and each of the raters in the study.

Residual variance component. The residual variance component was generated based on the information from one person, one item, one rater and other sources of error. The residual should theoretically be small compared to all of the other variance components in a G-Study. Unfortunately, in this study, the residual variance component accounts for the largest portion of variance: 0.30 and 35.8% of the total variance. This residual value means that after accounting for the variation in the main effects and multiple two-way interactions, 36% of the variance remains unaccounted for. This result would be more worrisome if the design had not been fully crossed. In a less than fully crossed design the two way interactions would become sources of error, depending on the nature of the lack of complete data.

G-Facets analysis. The results from this study produced a G coefficient of 0.86, which indicates the amount of variance associated with each applicant's score based on the universal score. The Swiss Society for Research in Education Working Group (2010) suggests that an acceptable G coefficient is one that is greater than or equal to 0.80. According to these standards, this study produced a G coefficient that adequately supports the precision of the measures produced. The relative G coefficient was used as opposed to the absolute G coefficient because the behaviour of each individual applicant is viewed in relation to the behaviour of all the other applicants. The G-facets analysis showed relative G coefficient values that ranged from 0.81 to 0.85 for individual items and 0.79 to 0.86 for

individual raters. Table 7 presents the G coefficient values associated with the G-Facets analysis for each individual rater and item.

Table 7

G-Facets Analysis

Facet	Level	Coef_G rel.	Coef_G abs.
Raters	Faculty Counsellor (New)	0.836198	0.830243
	Faculty Counsellor	0.790445	0.779478
	Faculty Non-Counsellor	0.858087	0.848327
	Student Counsellor I	0.817899	0.809834
-	Student Counsellor II	0.852909	0.849115
Items	Degree	0.841459	0.828426
	Writing Ability	0.834080	0.819554
	Fit of Goals	0.830586	0.814528
	Work Experience	0.815632	0.801002
	R1:Suitability	0.840492	0.833948
	R1:Quality	0.832698	0.817089
	R2:Suitability	0.836034	0.829761
	R2:Quality	0.813943	0.798169
	R3:Suitability	0.845146	0.835270
	R3:Quality	0.816160	0.800983

Given that the Rasch analysis revealed interesting results when the applicants were divided according to full-time or part-time studies, separate G-Studies for full-time and part-time applicants were conducted. The G-Studies for the full-time students are featured in Table 8.

Source	Variance Component	%	df	SE
Persons (P)	0.120499	13.5	36	0.031017
Raters (R)	0.011866	1.3	4	0.008122
Items (I)	0.122905	13.8	9	0.066886
P*R	0.070837	7.9	144	0.008364
P * I	0.222118	24.9	324	0.021929
R*I	0.062909	7.0	36	0.016184
P*R*I	0.282136	31.6	1296	0.011075
Coef_G relative	0.89			
Coef_G absolute	0.88			

Estimated G-Study Variance Components for Full-Time Applicants

The G-Studies summary for the part-time students is featured in Table 9.

Table 9

Estimated G-Study	Variance	<i>Components</i>	for Part-T	ime Applicants
		1 .	,	11

	Components							
Source	Variance Component	%	df	SE				
Persons (P)	0.013894	2.1	11	0.012038				
Raters (R)	0.005114	0.8	4	0.006905				
ltems (I)	0.081811	12.1	9	0.051668				
P*R	0.073803	10.9	44	0.015521				
P * I	0.155859	23.1	99	0.030207				
R*I	0.060311	8.9	36	0.019347				
P*R*I	0.284411	42.1	396	0.020161				
Coef_G relative	0.48							
Coef_G absolute	0.47							

When the applicants were divided according to whether they requested full-time or part-time program status, a G coefficient of 0.89 was observed for the full-time applicants and a G coefficient of 0.48 was observed for the part-time applicants. These results demonstrate that the applicants who applied for full-time studies were viewed as moderately heterogeneous and the applicants who applied for part-time studies viewed as homogeneous. Notice that the

full-time applicants produced a variance component of 0.120499 (13.5%) while the parttime applicants produced a variance component of 0.013894 (2.1%). As mentioned previously, homogeneity presents a problem in Generalizability analyses because the goal of G-theory is to describe the reliability of generalizing from a person's observed score to a universe of scores. If the sample selected for a G-Study is not representative of the population, then it becomes difficult to generalize the results back to the population. The other consideration that arises from separating the applicants according to program status is that there were only 12 applicants that applied for part-time studies. Without a suitable number of persons in the sample, G-Theory is not able to produce a G coefficient that supports the precision of the measures.

Decision studies. D-Studies were conducted as part of this study. D-Studies use data from a G-Study to provide information about the optimal conditions for future research designs (Shavelson & Webb, 1991). The results from the D-Studies reflect results for raters when items are fixed at ten and for items when the raters are fixed at five. Table 10 presents the D-Studies for the desired number of raters. Table 11 presents the D-Studies for the desired number of items.

D-Studies for Raters

	G-study	Option 1	Option 2	Option 3	Option 4	Option 5
	Level	Level	Level	Level	Level	Level
P	49	49	49	49	49	49
R	5	1	2	3	4	6
1	10	10	10	10	10	10
Rel. Err. Var. Coef_G	0.015501	0.077503	0.038751	0.025834	0.019376	0.012917
rel.	0.860690	0.552703	0.711924	0.787548	0.831724	0.881149
Rounded Abs. Err.	0.86	0.55	0.71	0.79	0.83	0.88
Var. Coef_G abs.	0.016355	0.081776	0.040888	0.027259	0.020444	0.013629
	0.854130	0.539401	0.700793	0.778431	0.824078	0.875413
Rounded	0.85	0.54	0.70	0.78	0.82	0.88

Table 11

D-Studies for Items

	G-study	Option 1	Option 2	Option 3	Option 4	Option 5
	Level	Level	Level	Level	Level	Level
P	49	49	49	49	49	49
R	5	5	5	5	5	5
I	10	6	8	12	14	16
Rel. Err. Var. Coef_G	0.031804	0.053006	0.039755	0.026503	0.022717	0.019877
rel.	0.714167	0.59986	0.666537	0.749891	0.777677	0.799907
Rounded Abs. Err.	0.71	0.6	0.67	0.75	0.78	0.8
Var.	0.040026	0.066711	0.050033	0.033355	0.028590	0.025016
Coet_G abs.	0.665022	0.543621	0.613633	0.704345	0.735406	0.760561
Rounded	0.67	0.54	0.61	0.7	0.74	0.76

The results of the D studies revealed that the most desirable measurement condition for the applicant evaluation process would be obtained by using four raters and ten items, although

it could be argued that three raters and ten items would also be acceptable. By keeping the number of raters fixed and varying the number of items, it would take sixteen items to reach an adequate measurement condition. The application selection committee has previously established the criteria they feel would predict individuals who would be best suited to the MEd-Counselling program; therefore, adjusting the number of items would be redundant and impractical in this testing situation. Figure 2 shows the results for each of the D-Studies in relation to one another.



Figure 2. Alternative D-Studies for determining the optimal number of raters and items.

The graphic representation of the D-Studies that examined all of the most feasible options for the desirable number of raters and items in this study suggests that a larger G coefficient is produced when the number of raters is adjusted more than when the number of items is adjusted. To achieve a G coefficient that is approximately 0.80, either a minimum of three raters or a minimum of sixteen items is required. As stated earlier, adjusting the number of items would not be an ideal option because adding six irrelevant items would decrease the probability that all of the items would be measuring one unidimensional construct, not to mention that five raters would still be required. Conversely, it would be reasonable to adjust the number of raters, especially in this case, considering that it would involve removing one or possibly two raters and still having only ten items.

Part-time Applicants Revisited

In the earlier discussion of the small sample size generated from the G-Studies that examined the part-time applicants, the researcher asked if any of the original five raters would be willing to evaluate the Northwest regional applicants. Four out of the five original raters agreed: Faculty Counsellor (New), Faculty Non-Counsellor, Student Counsellor I, and Student Counsellor II. The reason they were not initially rated with the Prince George campus applicants was twofold. First, the Northwest applicants were not competing with the Prince George applicants for seats. Second, the Northwest applicants were not rated because it was a non-competitive process in that every applicant that applied was offered a seat in the program as a part-time student. Nevertheless, the part-time applicants at the Prince George campus and the part-time applicants at the Northwest campus should be similar; therefore, the two samples were combined together to produce a part-time sample of 33 applicants. Rasch and G-Theory analyses were both conducted on the combined part-time applicant sample. The G-Studies summary for the combined part-time applicants is featured in Table

12.

Table 12

Estimated G-Study Variance Components for Combined Part-Time Applicants

	Components								
Source	Variance Component	%	df	SE					
Persons (P)	0.083924	9.4	32	0.024424					
Raters (R)	0.017857	2.0	3	0.012625					
ltems (I)	0.047478	5.3	9	0.035126					
P*R	0.061512	6.9	96	0.008883					
P*I	0.341691	38.5	288	0.034140					
R*I	0.065597	7.4	27	0.019381					
P*R*I	0.270177	30.4	864	0.012984					
Coef_G relative	0.85								
Coef_G absolute	0.81								

When the applicants from the Prince George campus and the Northwest campus were combined based on part-time program status, a G coefficient of 0.85 was produced. Returning to Table 8 where a G coefficient of 0.89 was produced for a sample of 37 full-time applicants, the results from the full-time and part-time split are now comparable with each other. The combined part-time applicants produced a variance component of 0.083924 (9.4%) for persons, which is much better than the variance component of 0.013894 (2.1%) for persons when there was only a sample size of 12. The results from Table 12 demonstrate that when examined together, all the applicants who applied in 2010 for part-time studies either in Prince George or in the Northwest were moderately heterogeneous.

Rasch revisited. Further results for the combined part-time applicants are found with the Rasch analysis rater and item reports. The Rasch summary for the items measurement report for the combined part-time applicants is depicted in Table 13.

Obsvđ	Fair		Model	Infit	Outfit			
Average	Average	Measure	S.E.	MnSq ZStd	MnSq ZStd	Nu Items		
3.7	2.65	.73	.11	0.7 -3	0.7 -3	1 Degree		
3.7	2.62	.76	.11	0.9 0	0.9 0	2 Writing Ability		
4.3	3.32	33	.12	0.9 0	0.9 0	3 Fit of Goals		
4.5	3.73	-1.05	.14	1.4 3	1.3 2	4 Work Experience		
4.3	3.43	52	.13	1.1 1	1.2 1	5 R1:Suitability		
4.1	3.07	.09	.12	1.1 0	1.1 0	6 R1:Quality		
4.1	3.18	09	.12	1.2 2	1.3 2	7 R2:Suitability		
4.0	2.98	.23	.12	0.9 0	0.9 0	8 R2:Quality		
4.1	3.17	08	.12	1.1 1	1.1 1	9 R3:Suitability		
4.0	2.95	.27	.12	0.8 -1	0.8 -1	10 R3:Quality		
Adj S.D51 Separation 4.18 Reliability .95								
Fixed (all same) chi-square: 173.1 d.f.: 9 significance: .00								
Random (normal) chi-square: 9.0 d.f.: 8 significance: .35								

Combined Part-time Applicants Items Measurement Report

In regard to the Rasch analysis, increasing the sample size of the part-time applicants produced results for items that were similar to the results for items when the part-time sample was 12. The work experience item was still considered to be the least difficult item (logit score = -1.05) and meeting the degree requirements (logit score = 0.73) and writing ability (logit score = 0.76) still appeared to be the most difficult items for part-time students. The reliability coefficient increased to 0.93 from 0.95, affirming that each of the ten items used to evaluate applicants varied in their level of difficulty. All of the items still have Infit and Outfit values that fall within the 0.5 to 1.5 range, which support unidimensionality and model fit. Based on these results it appears as though the Rasch model was not as sensitive to sample sizes as G-Theory. The Rasch summary for the rater measurement report for the combined part-time applicants is depicted in Table 14.

Obsvđ	Fair		Model	Ini	it	Outfi	lt		
Average	Average	Measure	S.E.	MnSq	zstd	MnSq	ZStđ	Nu	Rater
4.2	3.28	26	.08	1.3	4	1.4	4	1	Faculty Counsellor (New)
4.1	3.18	10	.13	0.8	-1	0.8	-1	2	Faculty Counsellor
4.2	3.23	18	.08	1.0	0	1.0	0	3	Faculty Non-Counsellor
4.0	3.01	.19	.08	0.8	-2	0.8	-2	4	Student Counsellor I
3.9	2.90	.35	.08	1.0	0	1.0	0	5	Student Counsellor II
Adj S.D.	.22 Sej	paration	2.33 Re	liabili	ity .8	4			
Fixed (all same) chi-square: 42.0 d.f.: 4 significance: .00									
Random (normal) chi-square: 4.1 d.f.: 3 significance: .25									

Combined Part-time Applicants Rater Measurement Report

The results for the combined part-time applicants produced some interesting findings. The most severe rater (R5) had a measure of 0.35 and the most lenient rater (R1) had a measure of -0.26. Notice by returning to Table 4 that the student raters (R4 and R5) were seen as the two most lenient raters. The findings shown in Table 14 produced the opposite results, the student raters (R4 and R5) are now seen as the two most severe raters (logit scores = 0.19 and 0.35). These results suggest that the faculty raters were more severe in their evaluation of the full-time applicants and more lenient with their evaluation of the part-time applicants. Accordingly, these results suggest that the students were more severe in their evaluation of the part-time applicants and more lenient with their evaluation of the full-time applicants. Myford and Wolfe (2004a) would suggest that the severity of the students' evaluations occured because they can better align themselves with the full-time applicants.

Full-time and Part-time Applicants Revisited

The results section of this paper began with a Wright variable map for relationships among facets for Prince George applicants (Figure 1). Given that incorporating the Prince George part-time applicants together with the Northwest part-time applicants seemed to balance out the research design, it seemed logical to conduct an analysis on all of the applicants, full-time and part-time from both the Prince George and Northwest campuses. The data are shown in Figure 3.

	Measr	+Student	-Rater				-Ite	ms			S.1	
+	3 +		+				+			-	+(5)	+
1												
							1					
		*					1					
		*										
I		*										
	1											
		* *										1
+	2 +	. ** .	+			-	+				+	+
1		* *					1				1	1
	I	* *									[
		* * * *					1					
	I	* * *									1	
		*******									ł	
		* * * * *										
		* * *										
	1	*****									4	
+	- 1 -	. ********	+				+				+	+
		* * * *										
		* * * * *										
		*										
		* * *					I					
		* *					4					
		* *										
							1	2	8	10		
							l					
		*					6					
*	0 *	* *	* R1 R2	R3	R4	R5 -	*				*	*
	l										3	
	l	*					3					
	I						7	9				
	I		ļ									
	I						5					
	I		ļ									
	ł		ł								ł	1
+	-1+		+			-	+			-	+(1)	+

Figure 3. Wright variable map for relationships among facets for Prince George and Northwest applicants.

The results of the many-facet Rasch analysis for all of the Prince George and Northwest applicants together are shown in Figure 3. The second column, "Student," describes the distribution of all 70 applicants. Consistent with the results generated in Figure 1, most of the applicants were situated within the 0 to 2 region on the logit scale indicating that they were proficient applicants. The fifth column, representing the item difficulties, was also consistent with the results generated in Figure 1. This means that difficult items like work experience and degree requirements were still challenging, and easy items like finding suitable individuals to provide references were still simple. The fourth column, titled "Rater," brought about interesting results: all of the raters were positioned at the 0 logit mark. Given that it is difficult to tell the exact variability of the raters on the Wright variable map, a combined Prince George and Northwest applicants rater measurement report was produced and is shown in Table 15.

Table 15

Obsvđ	Fair		Model	Infit	Outfit	
Average	Average	Measure	S.E.	MnSq ZStd	MnSq ZStd	Nu Rater
4.0	3.21	02	.05	1.2 2	1.2 3	1 Faculty Counsellor (New)
3.9	3.18	.01	.06	0.8 -3	0.8 -3	2 Faculty Counsellor
4.0	3.17	.02	.05	1.0 0	1.0 0	3 Faculty Non-Counsellor
4.0	3.17	.02	.05	0.8 -3	0.8 -3	4 Student Counsellor I
4.0	3.21	02	.05	1.2 3	1.2 2	5 Student Counsellor II
Adj S.D.	.00 Sej	paration	.00 Re	liability .0	0	
Fixed (a)	L1 same) cl	hi-square:	.8 d.f	.: 4 signifi	cance: .94	

Combined Prince George and Northwest Applicants Rater Measurement Report

The beginning of this section (under the subheading "Raters") stated that it would be ideal if the reliability coefficient was low. These results suggest two things. First, there were no significant differences between the five raters across the combined Prince George-Northwest sample. Second, the raters can actually be interchanged as their rating behaviour does not differ enough to be worrisome.

Probability

The Rasch analysis includes a report of probability curves that looks at how well each of the five categories of the rating scale functions. The probability curves for the 5-point scale used in this study are presented in Figure 4.

Probab	ility Cu	irves						
	3.0	-2.0	-1.0	0.0		0	2.0	3.0
	++	+		+	+-		+	++
1								
	1							
	1							1
	1111							
	11	1						1
Р		11						555
r	{	11						55
0	1	1					55	
b	l	11					55	Į
a		11					55	ļ
b			1		444	4444	55	
i			1	33333333	4444	444*	*	1
1	1		11 333	3*	*4	55	444	
i	1		133	44	333	55	444	1
t			3311	44	33	55	44	4
У	1	2222	2**2222*22	2 44	3**	*		444
		22222 3	3 11	***	55	33		4
	2222	2 333	44	11 2222	555	333		
	222	333	444	111 2**	*5	33	333	
	3333	333 4	44444	555***11	2222222		333333	3
0	******	********5	555555555555555555555555555555555555555	55	11111111	******	*******	****
	++	+		+	+-		+	++
-	3.0	-2.0	-1.0	0.0	1.(00	2.0	3.0

Figure 4. Probability curves for the 5-point rating scale used to evaluate MEd-Counselling applicants from both the Prince George and Northwest campuses.

The probability curves showed some overlap and disordering in the steps between categories one, two, and three. Linacre (2010) would recommend asking whether the categories are different enough to merit a separate categorical point on the rating scale. The lower

threshold of the second probability curve indicated that the raters were unable to clearly distinguish between those rating categories. Perhaps the second point on the rating scale should be collapsed into either the first point or the third point category to create more distinctive boundaries between each of the categories. Examination of the mode in the probability curves figure (Figure 4) showed no separation between category one and category three. In this case, looking at the median or mean thresholds provided a better interpretation about where the second category is aligned on the logit scale in relation to the other categories. Figures 5 provided a scale structure for each of the five categories used to scale and rate all of the applicants.

				<u> </u>			
Scale struct	ure						
Measr:-3.0	-2.0	-1.0	0.0	1.0	2.0	3.0	
+	+	+	+	+	+	+	
Mode:<1	(^)	13	^34	^	-45	(^)>	
Median:<1	(^):	L2^23	^34	^	45	(^)>	
Mean:<1	(^) 12	23	^	^	45	5>	
+	+	+	+	+	+	+	
Measr:-3.0	-2.0	-1.0	0.0	1.0	2.0	3.0	_

Figure 5. Scale structure for applicants from Prince George and Northwest campuses.

G-Theory revisited. Unfortunately, G-Theory was unable to supply any information about the categories used in the rating scale. However, a G-Theory analysis was conducted for the combined Prince George and Northwest applicants and the G-Studies summary for the combined Prince George and Northwest applicants is featured in Table 16.

	Components							
Source	Variance Component	%	df	SE				
Persons (P)	0.052949	3.0	69	0.037410				
Raters (R)	0.284887	16.1	4	0.170891				
ltems (I)	0.031622	1.8	9	0.020051				
P*R	0.776699	43.8	276	0.065885				
P*I	0.280245	15.8	621	0.019512				
R*I	0.034006	1.9	36	0.008830				
P*R*I	0.313438	17.7	2484	0.008890				
Coef_G relative	0.25							
Coef_G absolute	0.20							

Estimated G-Study Variance Components for Combined Prince George and Northwest Applicants

The G coefficient warrants special consideration here given that this G-Study has produced a value of 0.25. One may remember that R2: Faculty Counsellor did not rate the Northwest part-time applicants. G-Theory does not handle the issue of missing data well. In this particular case, the EDUG software treated R2's missing data as "0" in the data set; the "0" in the data set lead to an analysis of what appeared to be a 6-point scale to which all of the part-time Northwest applicants were given exactly the same ratings of "0" according to R2's ratings. The variance component for the main effect of raters was 0.284887 and accounted for 16.1% of the total variance, which was by far the largest amount of variance seen among raters in any of the G-Studies conducted with these raters. The Rasch analysis seen earlier (Figure 3) used the same missing data set as this analysis and it was not problematic. Rasch detected the missing data and generated an output based on the estimates of the data that R2 previously rated.

Chapter Five: Discussion and Conclusions

The primary goal of this study was to assess the overall effectiveness of a graduate student applicant selection process as it currently exists at UNBC. This process included an analysis of the items, raters, and applicants. The Many-Facet Rasch Model and Generalizability Theory were chosen because of their ability to provide relevant and credible information about rater, item and applicant consistencies. The Rasch analysis was able to show that all of the items fit within a unidimensional construct. It also informed the researcher that the rating behaviour of the participants was acceptable by providing measures that reflected the severity or leniency of each of the raters in this study. Conversely, the G-Theory analysis was able to account for the proportion of variance that each of the facets contributed. Furthermore, G-Theory provided information about alternative research designs that would be best employed in the future. Using both the Many-Facet Rasch Model and Generalizability Theory showed that the two methodologies complement each other well in their abilities to describe the elements of variability within data. Both methodologies provided the researcher with information about item, rater, and applicant characteristics in a way that could be used to make inferences about the findings of this study.

The secondary goal of this study was to investigate the variability of the second year MEd-Counselling students who were acting as raters alongside the faculty raters. When the sample size was adequate, the results indicated that the student raters behaved no differently than the faculty raters. A complete applicant report for all 70 applicants from both the Prince George and Northwest campuses can be found in Appendix E.

58

Items. By conducting the Rasch analysis, the researcher showed that all of the items differed from each other in degree of endorsement (difficulty); however, they were not different enought that the items could not be measured as a single unidimensional construct. In analyzing the items used to rate applicants applying to the MEd-Counselling program, the Rasch model conveyed the range of each item's level of difficulty. The Rasch analysis provided precise information about the data. For instance, the Rasch analysis indicated to what degree the work experience item was potentially misfitting by displaying large Mean square Fit values (Infit = 1.5; t = 4 and Outfit = 1.5; t = 4). These values were later discovered to be caused by the combined full-time and part-time samples of applicants. The Rasch analysis was also able to show that the relevant degree item was too predictable by the low Infit and Outfit scores that were produced (Infit = 0.6; t = -4 and Outfit = 0.6; t = -4). Based on this information it would be worthwhile to remove this item from the rating scale and have it coded by an administration assistant. An assistant would likely give the same value as the application selection committee. The FACETS program (Linacre, 1996) was able to provide Fit statistics (Infit and Outfit values) for each item, rater, and applicant involved in this study. The Rasch analysis produced a separation ratio (5.67) and reliability index (0.97) for the items; these statistics mean that the items were highly separable, but also that the differences between the items was over five times greater than the error associated with the measurement model.

The G-Theory analysis indicated specific variance components for each of the facets and possible combination of interactions identified in this study. For example, the G-Theory analysis demonstrated that the items differ slightly in difficulty by showing that 7.9% of the variance was due to the items facet. The G-Theory analysis produced individual G

coefficients for each of the ten items used to measure the applicants, both by relative and absolute standards. In conducting a Rasch analysis and a G-Theory analysis for the full-time and part-time applicants, the researcher was able to show a seeming invariance for items across each sample of applicants drawn from larger populations.

Raters. There were five raters in total who participated in this study: three faculty members and two students. As far as the rater analysis was concerned, the Rasch analysis was most informative in providing information about how each rater behaved individually because of its ability to transform the raw scores into logit scores and place them on an interval scale of measurement. The rater measurement report shown in Table 4 produced a 0.31 logit spread among the five raters. A 0.31 logit spread is low, considering the diversity of knowledge and experience among all five raters. The study conducted by Sudweeks and colleagues (2005) produced a logit spead of 0.51 among raters. The Rasch analysis showed that the most severe rater (0.15 logits) was a new faculty member who had a counselling background, but no previous experience with this task at this institution. The next severe rater (0.09 logits) was the other faculty member, one who did not have a counselling background, but had considerable experience with this process. The faculty member who had a counselling background and was familiar with the application process was situated in the middle of the five raters (0.01 logits). Overall, the two student raters were the most lenient of all the raters (-0.09 and -0.16 logits), with the first being overly constrained and the other being somewhat erratic according to t statistics (Infit = -2 and 3; Outfit = 4 and 3). Verbal communication between the two student raters revealed that, although they have limited experience in the field of counselling, neither of them felt qualified enough to be rating the applicants on items concerning reference suitability and work experience. An

interesting finding related to the behaviour of the raters was discovered when analyses were conducted separately based on the full-time and part-time status. The students rated the parttime applicants more severely than the full-time applicants. Meanwhile, the faculty members rated the full-time applicants more severely than the part-time applicants. Perhaps, as Myford and Wolfe (2004a) suggested, these effects occurred because the raters (full-time students) were able to identify in some way with the applicants whom they were rating. Further investigation of this issue, as revealed by one of the student raters, suggested that although work experience is essential in the field of counselling, a strong level of writing ability is necessary in order to meet the demands of the UNBC MEd-Counselling program. Personal communication between the researcher and the faculty member without a counselling background resulted in the faculty member asserting that part-time applicants bring a wealth of knowledge to the program, having worked in the helping profession for an extended amount of time. The student rater who believed that a strong level of writing ability is critical admitted that the work experience component is a crucial one; but the rater argued, that once admitted into the program, both the full-time and the part-time students are required to perform at the same level academically, given that each student is working towards fulfillment of the requirements for a Master of Education degree. Nevertheless, the results generated from the sample of combined Prince George and Northwest applicants suggested that the variation in rating behaviour may not have been just a difference of opinion, but rather a result of a small sample size for the part-time applicants. When the sample sizes for the full-time and part-time applicants were above 30, no significant differences were found between any of the raters (Table 15). The three faculty raters had logit scores of -0.02, 0.01, and 0.02 that produced a 0.04 logit spread. This result is
comparable with the study conducted by Smith and Kulikowich (2004) that produced a 0.04 logit spread, but for only two raters. The two student raters had logit scores identical to two of the faculty raters, which produced an overall 0.04 logit spread for all the raters. The five raters behaved in a way that made it difficult to distinguish between each of them, suggesting that this process may only require one or two raters to evaluate and select the successful applicants.

The Generalizability analysis suggested that all of the raters used the rating scale similarly (0.5%) and the relatively small (4.9%) rater-by-item interaction supports the use of a less than fully crossed rater by applicant design. However, the 9.2% variance component for the rater-by-applicant interaction indicated otherwise. This was not seen as an issue when the full-time and part-time applicants were analyzed separately (Table 8 and Table 12). This claim is further supported by the Rasch analysis results. Likewise, a large applicant-by-item interaction (30.4%) yielded similar variance component values for the analysis of the two separate populations. However, the applicant-by-item interactions for full-time (24.9%) and for part-time (38.5%) were still larger than a researcher would hope for. At any rate, as all applicants needed to respond to all items, this interaction does not act as a source of error in this design.

The G-Facets analysis (Table 7) provided relative G coefficients for each of the raters, which indicated how accurate each rater's scoring behaviour was in relation to universal scoring behaviour. Among the decision studies made throughout this analysis, the ideal condition would be four raters across ten items. This design would produce a G coefficient of 0.83 and a relative error variance of 0.019. The G-Study analysis reinforced the Rasch results of two populations of applicants existing within the sample that applied to the MEd-

Counselling program: full-time and part-time based on revealed sources of variation. Given that G-Theory looks at how well a single person's score can be generalized across the universe of scores, it is important to have a high percentage of variance accounted for by persons. Larger variance components for persons indicate that the sample is more heterogeneous, and thus more representative of the universal population.

In retrospect, the more informative data came from conducting both the Rasch and G-Theory analyses, then presenting the information to each of the raters, and asking them why they felt they rated the applicants on the items the way that they did. Engaging in personal communication with each of the raters enhanced the quantitative data by adding the unique qualitative perspective of each of the raters.

Applicants. From the perspective of a researcher and counselling practitioner, it is more justified to view all of the applicants together even though the Rasch and G-Theory analyses revealed the presence of two distinct applicant populations. The separation of 2.79 and the reliability of 0.89 for the Prince George applicants, as well as the separation of 2.57 and the reliability of 0.87 for the Prince George and Northwest applicants, indicates that there was relatively strong person heterogeneity among the applicants. When all of the applicants from both the Prince George and Northwest campuses were considered as one population there were only three outliers: applicant 27 (Infit = 1.70, t = 2; Outfit = 1.60, t = 2) who applied for full-time studies, applicant 48 (Infit = 1.60, t = 2; Outfit = 1.60, t = 2; Outfit = 1.70, t = 2) who applied for part-time studies in Prince George, and applicant 66 (Infit = 1.60, t = 2; Outfit = 1.70, t = 2) who applied for part-time studies in the Northwest. Even though each population presents different aspects, the data from the Rasch analysis suggests that how the applicants are being rated on the items is measurable as a single unidimensional construct. This is ideal

considering that, once offered a seat in the program, all successful applicants will be working towards completing the same degree requirements regardless of their full-time or part-time status. It was still useful to examine these two populations independently of one another because they are non-competitive with each other. The university has a specific number of seats in the program available for full-time students and a specific number of seats available in the program for part-time students; therefore, even though all of the applicants were assessed together, competition for letters of acceptance was based on whether the applicant indicated the intention of full-time or part-time study.

Student Raters. Based on the information presented in this study, there is a case to be made for permanently incorporating student raters as part of the application selection process. The two student raters, who completed all course requirements and practicum, brought an alternative perspective to the application selection committee. Corey, Corey, and Callanan (2007) state in relation to professional competence and training that:

A number of programs have both faculty members and graduate students on the reviewing committee. If many sources are considered and if more than one person makes a decision about whom to select for training, there is less likelihood that people will be screened out on the basis of the personal bias of one individual. (p.321)

The application selection committee may consider having a student on the application selection committee to lighten the workload of faculty members during such a busy time of the semester. There was no strong evidence that the student raters behaved in such a way that would be worrisome. If the committee would like to feel more confident and secure in assessing the ability of student raters, then the report of unexpected responses (see Appendix F) produced by the FACETS program (Linacre, 1996) could be examined. When the fulltime and part-time applicants from both the Prince George and Northwest campuses were analyzed together, there was actually no difference between the rating behaviour of the students and the rating behaviour of the faculty members. It is certainly a viable option to have second year counselling students acting as raters alongside the faculty raters. As a matter of fact, the Social Work hiring committee at UNBC has an undergraduate student who sits on the committee and has the same level of influence as any one of the other members on the committee.

Conclusions

After exploring the relationship between the Many-Facet Rasch Model and Generalizability Theory, it appears that each methodology has its prevailing strengths and weaknesses. The strengths of the Rasch model included greater detail when focusing on the individual elements of each facet and supplying error indicators for each element as well as a remarkable ability to handle small sample sizes and missing data. These strengths suggest that the Rasch model is robust to violations that many other models are unable to withstand. Some of the weaknesses of the Rasch model relate to its simplicity. The Rasch model is not overly complicated, which has some researchers convinced that it is not a viable model. Also, the lack of concrete rules relating to things such as sample size and Fit statistics has been a documented source of frustration for researchers. The strengths of the G-Theory included the ability to provide variance components for each facet's main effect and all possible interactions, the freedom to make relative or absolute decisions, and the decision studies feature that displays reliability measures for various designs. Some of the weaknesses of G-Theory have to do with its inability to compensate for small sample sizes and missing data. When it comes to the preference for one methodology over the other, the research questions should guide the preferred approach used for the analysis.

In asking whether the applicant selection process at UNBC produced a unitary score that can be used to rank all individuals applying to the counselling program, the Rasch analysis triumphed over G-Theory because it was able to produce Fit statistics for each individual item.

For the question of whether the 5-point rating scale used to evaluate the applicants served as an appropriate measurement tool, the Rasch and G-Theory analyses were both able to generate satisfactory results. The Rasch analysis produced Fit statistics, severity measures, a separation ratio, reliability score, and probability curves that provided information about the rating scale's performance. The G-Theory analysis generated variance components for items, applicants-by-items, and raters-by-items. The 8% variance accounted for by the main effect of items further supports the 5-point rating scale used to evaluate the applicants.

In considering the rating characteristics of the participants chosen to rate on the selection committee, both the Rasch and G-Theory analyses suggested that the raters were suitable as groups (0.31 logit spread, 0.73 reliability index, and p < .005; 0.5% variance for rater main effect, 9.2% variance for applicant-by-rater interaction, and 4.9% for rater-by-item interaction). However, only the Rasch analysis provided information about how the raters behaved as individuals (Fit statistics and severity measures for each rater).

Finally, the Rasch model is a viable method for dealing with rater differences because of its ability to produce severity measures, observed averages and fair averages for each rater. In conclusion, based on the nature of this study, the Rasch analysis seemed to be more advantageous than G-Theory. This advantage comes from Rasch analysis' ability to perform logarithmic transformations of ordinal raw data to interval measures, and its ability to produce Fit statistics for individual items, raters, and applicants that alert the researcher to

possible violations within the data. The greatest benefit of using Rasch analyses, which became even more apparent as this study progressed, was the Rasch model's ability to handle large amounts of missing data and relatively small sample sizes.

Limitations of the Design

One limitation that is unique to this particular study is that the counselling coordinator, who regularly holds this position and serves on the selection committee, is on sabbatical this year. This absence means that the selection committee for the September 2011 intake will have no data on this particular rater unless it could be taken from a previous year when she chaired the selection committee.

The large residual variance component of 35.8% for the 49 applicants from the Prince George campus is a source of concern, considering the design and the amount of information that was accumulated by simultaneously employing two different measurement models. The researcher's ability to employ a fully crossed design was beneficial in explaining the findings from the Generalizability analysis; however, a fully crossed design may not always be a realistic option in the future.

The gender dynamic consisted of two males and three females. There were three faculty raters and two student raters on this application selection committee. Given that four out of the five raters have a counselling background, further analysis regarding gender, age, academic status, and level of counselling experience would have made this study more informative.

Recommendations for the Application Selection Committee

Since the results of the analysis appeared to indicate some variation in the selection of applicants that would be best suited to pursue a Masters of Education in Counselling degree, the research findings warrant addressing and possibly adding the following components.

Faculty agreement. Based on the Fit statistics (Infit = 1.50; Outfit = 1.50) and G coefficient (0.816) values for the work experience item, it would be worthwhile to explore other steps that can be taken to ensure that this item stays constant across different samples of applicants applying to the MEd-Counselling program. One consideration would be to create a category that looks at the fit of prospective students to Education faculty members. The term "fit" is used here to imply that the applicant has provided evidence that they would potentially have the same opinion as some of the current faculty members in regard to theoretical approach or research interests.

Northern perspective. UNBC is interested in training counsellors who have a passion for their work; especially those who want to work in the North. Another option for the committee is to create a Northern experience item that would allow members on the application selection committee to assess an applicant's suitability and fit not only with the program, but with the university and the community that is established in Prince George. This would allow the raters the opportunity to judge the quality of an applicant's work and lived experience in the North, which might adjust the high Fit statistics (Infit = 1.50; Outfit = 1.50) and low G coefficient (0.816) values for the current work experience item.

Interviews. Through personal communication between the researcher and one of the faculty counsellors, another method of evaluating applicants would be through phone or inperson interviews. In the field of counselling where persona and aura play a vital part in the

therapeutic relationship, it would certainly benefit the School of Education to pre-screen applicants through interviewing. An interview would allow the raters to clarify any information that may have been ambiguous to the raters. This would likely sort out some of the issues with the work experience item and reduce the number of unexpected responses generated from the raters (see Appendix F for the complete list).

Applicant waitlist. Over the years, since the program first began, the number of applications the university receives from individuals wishing to enter the MEd-Counselling program has steadily increased. This year, the university received almost 50 applications from prospective students. According to the Wright variable map displayed in Figure 1, the applicants are outperforming the items. This was observed by the large number of students sitting above the 1.0 logit mark. These data are strong in the sense that the university was able to select among top quality applicants, but weak in the sense that there were a large number of quality applicants that were not offered a seat in the MEd-Counselling program. At this time there is no current waitlist policy for applicants that were not offered a letter of acceptance. Perhaps the Masters of Education program should look at drafting a protocol for those applicants who meet the criteria, but were not fortunate enough to be accepted into the program due to the relative competition.

Good measurement practice. The 4.9% rater-by-item interaction produced by the G-Theory analysis was low, but not unsubstantial. As part of good measurement practice, the names and other personal identifiers of the applicants were blanked out in an attempt to protect confidentiality and alleviate anything that could bias the effects of a particular rater. Continuing to engage in this process in the future might reasonably be assumed to help keep the rater-by-item interactions low. As mentioned previously, in the methods chapter, the researcher provided the raters with instructions for how to use the rating scale and information about the most commonly identified rater effects, biases, and errors as well as the consequences of committing such rating errors. This practice seemed to work well as the results of 0.5% variance accounted for by the main effect of raters with the G-Theory analysis and a 0.04 logit spread from the Rasch analysis demonstrated.

Recommendations for Future Research

In the future, assessing the MEd-Counselling program, using the same design and participants would be ideal. This would allow the researchers to examine rater drift (the rating patterns and behaviour over time). One of the most common facets analyzed with both Rasch analysis and Generalizability Theory is occasions. The data for this study was gathered on one occasion, which does not allow for the opportunity to examine item difficulty, rater behaviour, or applicant quality over any period of time. As mentioned previously, the MEd-Counselling coordinator is currently on sabbatical, which means there was no data suggesting where she fits with the other raters on the application selection committee. Replicating this study to include the MEd-Counselling coordinator next year could provide useful information about how she would have fit with the raters used in this study. The application selection committee could also investigate and experiment with other potential items, such as adding a supplementary item to settle the variability of the work experience item or removing the relevant degree item from the rating scale and having the relevant degree coded by one person.

Another recommendation for the future, given that the data now currently exists, is to try using Rasch analysis and Generalizability Theory in a design that is not fully crossed to see what degree of overlap is necessary to have raters review and score all of the application packages received each year with accuracy and precision.

The final recommendation, given that the data from this study has been made available to the School of Education, is to qualitatively and quantitatively examine the successful applicants to see what prompted them to apply to the MEd-Counselling program at UNBC. A great follow-up study to this one would be to collect data on how well each successful applicant performed in the program and compare it with the ranking they had when they entered the program.

In conclusion, both the Many-Facet Rasch Model and Generalizability Theory have strengths. Each methodology was designed with an idea of the optimal conditions that would warrant the use of that particular methodology. Research in the area of measurement requires researchers to make judgements as to whether the measurement context is appropriately suited to the methodology. Sometimes one methodology is not sufficient enough to adequately measure all of the questions that a researcher has. Therefore, with any analysis, it may be necessary to find two or more measurement models that can be combined to make the most out of the information situated within the data.

References

- Andrich, D. (1996). Measurement criteria for choosing among models with graded responses. In A. von Eye & C. C. Clogg (Eds.), *Categorical variables in developmental research: Methods for analysis.* (pp. 3-35). San Diego, CA: Academic Press.
- Atilgan, H. (2008). Using generalizability theory to assess the score reliability of the special ability selection examinations for music education programmes in higher education. *International Journal of Research & Method in Education*, 31(1), 63-76. doi: 10.1080/174372708011919925
- Bond, T. G., & Fox, C. M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chang, W. & Chan, C. (1995). Rasch analysis for outcome measures: Some methodological considerations. *Archives of Physical Medicine and Rehabilitation*, 76(1), 934-939.
- Corey, G., Corey, M. S., & Callanan, P. (2007). *Issues and ethics in the helping professions* (7th ed). Belmont, CA: Thomson Books/Cole.
- Engelhard, G. (1992). The measurement of writing ability with a many-facet Rasch model. *Applied Measurement in Education*, 5(3), 171-191.
- Engelhard, G. (1994). Examining rater errors in the assessment of written composition with a many-facet Rasch model. *Journal of Educational Measurement*, 31(2), 93-112.
- Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counselling psychology research. *Journal of Counseling Psychology*, 45(1), 30-45.
- Hurlburt, R. T. (2006). *Comprehending behavioral statistics* (4th ed). Belmont, CA: Thomson Wadsworth.
- Johnson, D. W., & Johnson, F. P. (2003). *Joining together: Group therapy and group skills* (8th ed). Boston, MA: Pearson Education.
- Kieffer, K. M. (1999). Why generalizability theory is essential and classical test theory is often inadequate. *Advances in Social Science Methodology*, 5(1), 149-170.
- Kim, S. C., & Wilson, M. (2010). A comparative analysis of the rating in performance assessment using generalizability theory and the many-facet Rasch model. In M. L. Garner, G. Engelhard, W. P. Fisher, & M. Wilson (Eds.), Advances in Rasch measurement volume 1 (pp. 304-327). Maple Grove, MN: JAM Press.
- Linacre, J. M. (1989). Many-facet Rasch measurement. Chicago: MESA.

- Linacre, J. M., Wright, B. D., & Lunz, M. E. (1990). A facets model for judgmental scoring. Retrieved from <u>http://www.Rasch.org/memo61.htm</u>.
- Linacre, J. M. (1995). Categorical misfit statistics. *Rasch Measurement Transactions*, 9(3), 450.
- Linacre, J.M. (1996). FACETS: A computer program for analysis of examinations with multiple facets, version 3.03. Chicago: MESA.
- Linacre, J. M. (1997). *Judging plans and facets*. Retrieved from <u>http://www.Rasch.org/rn3.htm</u>.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3(2), 103-122.
- Linacre, J. M., & Wright, B. D. (2004). Construction of Measures from many-facet data. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp.296-321). Maple Grove, MN: JAM Press.
- Linacre, J. M. (2010). *Rasch measurement: Core topics*. Retrieved from <u>http://courses.statistics.com/index.php3</u>.
- Liu, O. L., Minsky, J., Ling, G., & Kyllonen, P. (2009). Using the standardized letters of recommendation in selection: Results from a multidimensional Rasch model. *Educational and Psychological Measurement*, 69(3), 475-492. doi: 10.1177/0013164408322031
- Lochhead, L. (2009). Assessment of perceived functional capacity: Using Rasch analysis to evaluate the measurement properties of four perceived pain & disability scales. (Master's thesis). University of Northern British Columbia, Prince George.
- Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3(4), 331-345.
- Lunz, M. E. (1999). A longitudinal study of judge leniency. *Popular Measurement*, 47(1), 46-47.
- MacMillan, P. (2000a). Simultaneous measurement of reading growth, gender, and relative age effects: Many-faceted Rasch applied to CBM reading scores. *Journal of Applied Measurement*, 1(4), 393-408.
- MacMillan, P. D. (2000b). Classical, generalizability, and multifaceted Rasch detection of interrater variability in large, sparse data sets. *The Journal of Experimental Education*, 68(2), 167-190.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.

Matt, G. E. (2010) *Generalizability theory*. Retrieved from <u>http://www.psychology.sdsu.edu/faculty/matt/Pubs/GThml/GTheory_GEMatt.html</u>.

- McHorney, C. A., Haley, S. M., & Ware, J. E. (1997). Evaluation of the MOS SF-36 physical functioning scale (PF-10): II. Comparison of relative precision using likert and Rasch scoring methods. *Journal of Clinical Epidemiology*, 50(4), 451-461.
- Myford, C. M., & Wolfe, E. W. (2004a). Detecting and measuring rater effects using manyfacet Rasch measurement: Part I. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 460-517). Maple Grove, MN: JAM Press.
- Myford, C. M., & Wolfe, E. W. (2004b). Detecting and measuring rater effects using manyfacet Rasch measurement: Part II. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 518-574). Maple Grove, MN: JAM Press.
- O'Neill, T.R. (1999). Adjusting for rater severity over time. *Popular Measurement*, 47(1), 46-47.
- Oosterveld, P., & ten Cate, O. (2004). Generalizability of a study sample assessment procedure for entrance selection for medical school. *Medical Teacher*, *26*(7), 635-639. doi: 10.1080/01421590400004874
- Pedersen, G., Hagtvet, K. A., & Karterud, S. (2007). Generalizability studies of the global assessment of functioning: Split version. *Comprehensive Psychiatry*, 48(1), 88-94. doi: 10.1016/j.comppsych.2006.03.008
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, II: University of Chicago Press.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage Publications.
- Smith, E. V. (2004). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp.93-122). Maple Grove, MN: JAM Press.
- Smith, R. M. (2004). Fit analysis in latent trait measurement models. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp.73-92). Maple Grove, MN: JAM Press.

- Smith, E. V., & Kulikowich, J.M. (2004). An application of generalizability theory and many-facet Rasch measurement using a complex problem-solving skills assessment. *Educational and Psychological Measurement*, 64(4), 617-639. doi: 10.1177/0013164404263876
- Stewart, I. (2006). Letters to a young mathematician. New York, NY: Basic Books.
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9(1), 239-261. doi: 10.1016/j.asw.2004.11.001
- Swiss Society for Research in Education Working Group. (2010). *EDUG user guide, version* 6.0. Neuchatel, Switzerland: Edumetrics.
- Winne, P. H., Nesbit, J. C., Kumar, V., & Hadwin, A. F., Lajoie, S. P., Azevedo, R. A., & Perry, N. E. (2006). Supporting self regulated learning with g-study software: The learning kit project. *Technology, Instruction, Cognition and Learning*, 3(1), 105-113.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Wright, B. D. (1997). Fundamental measurement for outcome evaluation. *Physical Medicine* and *Rehabilitation*, 11(2), 261-288.

Appendix A: The Biblical Approach to Understanding Additive Measurement

This passage was taken from Ian Stewart's (2006) "Letters to a young mathematician"

"The flood has receded and the ark is safely aground atop Mount Ararat; Noah tells all the animals to go forth and multiply. Soon the land is teeming with every kind of living creature in abundance, except snakes. Noah wonders why. One morning two miserable snakes knock on the door of the ark with a complaint. 'You haven't cut down any trees.' Noah is puzzled, but does as they wish. Within a month, you can't walk a step without treading on baby snakes. With difficulty he tracks down the two parents. 'What was all that with the trees?' 'Ah,' says one of the snakes, 'you didn't notice which species we are.' Noah still looks blank. 'We're adders, and can only multiply using logs.'"

This joke is a multiple pun: you can multiply numbers by adding their logarithms.

Appendix B: Participant Consent Form "Pick Me, Pick Me, I Want to Be a Counsellor": Assessing MEd-Counselling Applicants Using Rasch Analysis and Generalizability Theory

You are invited to participate in the evaluation of the Masters of Education-Counselling program at the University of Northern British Columbia. This evaluation is directed by Stefanie Sebok, MEd Candidate, in collaboration with the School of Education at the University of Northern British Columbia. Stefanie is a student in the School of Education at the University of Northern British Columbia and you may contact her by phoning (250) 960-5671 if you have any questions.

The purpose of this research project is to provide information about the applicants who have applied to the MEd-Counselling program for intake in September 2010, to evaluate the effectiveness of the items used by application selection committee to score applicants, and to assess the rater characteristics of each member on the application selection committee. This evaluation will inform the School of Education and provide feedback to refine the selection process in the future.

All information we ask you to provide is confidential and is only seen by the researcher and her supervisor, Dr. Peter MacMillan. No names will appear on any outputs; numerical codes or coded initials will be used instead of names for presentation purposes. All information that you provide will be stored in a locked filing cabinet until the study is completed. Once this study has been completed, all application packages will be returned to the Chair of the application selection committee.

There are no known risks from participating in this research. Your participation in this study will help increase knowledge about ways to enhance the overall MEd-Counselling application selection process in the future.

Your participation in this study is voluntary. You are free not to participate and there are no negative consequences for not participating. If you decide to participate and then change your mind, you may withdraw at any time without any consequences or any explanation. If you withdraw from the study, your data will be removed from the analysis.

The Research Ethics Board (REB) at the University of Northern British Columbia has reviewed this study and granted permission to move forward as this study constitutes standard program evaluation. If you have any concerns as a participant in this study you may contact the Office of Research at UNBC (250) 960-5820, or raise any concerns you might have by contacting Stefanie Sebok or Dr. Peter MacMillan, Supervisor, at the University of Northern British Columbia at (250) 960-5828.

Your signature below indicates that you understand the above-noted conditions of participation in this study and that you have had the opportunity to have your questions answered by the researchers.

I ______(PRINT NAME) agree to participate in the evaluation of the MEd-Counselling application selection process.

Signature

Appendix C: Research Ethics Letter

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

RESEARCH ETHICS BOARD

MEMORANDUM

Stefanie Sebok
Peter MacMillan
Henry Harder, Chair Research Ethics Board
April 1, 2010
They were too Rasch with my application

Thank you for your application regarding the above noted project. The committee has discussed the information and is supportive of your involvement with the data analysis of and felt that the project falls outside the purview of the ethics committee as there are no human subject participants.

The committee appreciates the work you have done on this application and for allowing the UNBC REB to review your project and we are happy to perform our due diligence when a researcher becomes involved in a project.

Regards,

Henry Harder

Appendix D: Assessment Criteria

Guide for Assessing MEd in Counselling Applicants: Intake 2010

These criteria will serve as a guide to informing the discussion of applicants. The final decision pertaining to acceptance of applicants into the MEd Counselling specialization will be made by committee vote.

Grade Point Average (GPA) (maximum 4.33)

*GPA is part of the applicant's overall score; however, the raters do not affect the score given for GPA

Relevant educational degrees (maximum 5 points)

- Graduate degree in Psychology, Social Work, Child/Youth Care, Education 5 pts
- Relevant undergraduate degree in Psychology, Social Work, Child/Youth Care, Education
- or graduate degree in Nursing, Health Sciences, First Nations Studies, Criminology 4 pts
- Undergraduate degree in Nursing, Health Sciences, First Nations Studies, Criminology or graduate degree that has some relevance **3 pts**

• Undergraduate degree that has some relevance or graduate degree that has little opportunity for written expression – 2 pts

Undergraduate degree that has little opportunity for written expression – 1 pt

Statement of academic/research interests (5 + 5 = 10 points maximum)
Competence in writing - (1 pt = very poorly written, 2 pts = poorly written, 3 pts = acceptable, 4 pts = well written, 5 pts = very well written)

There is a fit between goals of applicant and the program – (1pt = not a very good fit, 2 pts = not a good fit, 3 pts = acceptable, 4 pts = a good fit, 5 pts = a very good fit)
* An application may be rejected if writing is not of suitable quality or goals are not compatible with the program

Relevant employment/volunteer work (maximum 5 pts)

- 5+ years of full time work experience in a helping profession 5 pts
- 1-5 years of full time work experience in a helping profession or 5+ years of part time work experience in a helping profession 4 pts
- Less than 1 year of full time work experience in a helping profession or 1-5 years part time work experience in a helping profession or 5+ years of volunteer experience in a helping profession **3 pts**
- 1-5 years volunteer experience in a helping profession or less than 1 year part time work experience in a helping profession 2 pts
- Less than 1 year volunteer experience in a helping profession 1 pt

References $(5 + 5 = 10 \text{ points each for an overall maximum of 30 points)$

• Referee's suitability for the Counselling program – (1 pt = very unsuitable, 2 pts = unsuitable, 3 pts = acceptable, 4 pts = suitable, 5 pts = very suitable)

* The best references are those from a university/college instructor, employment supervisor, referral agent that has observed the applicant in some extended capacity

- Quality of the reference based on professional judgment of relevant counselling qualities
- (1 pt = poor, 2 pts = satisfactory, 3 pts = good, 4 pts = very good, 5 pts = excellent)
- * An application may be rejected based on issues of serious concern in the references

Rater # ____

Applicant's Name	GPA	Degree	Statement	Employment	References	Total	Rank
			Interest/				
	(Max	(Max	Research	(Max 5)	(Max 10 pts		
	4.33)	5)			each for a		
					pts)		
			(Max 10)		_		
	C=		(112022 10)		Max 5 pts		
	conditional		Max 5 pts		for suitability		
					of each ref		
			for writing		Į		
					Mary 5 miles		
			competence		Iviax 5 pts		
					for quality of each ref		
			Max 5 pts				
			for fit of		Veto power		
			goals				
			Veto				
			power				
	·						
				·		·	
1			l i				

Note: Students who are required to submit TOEFL results may also be asked to participate in a telephone interview.

Appendix E: Report for Prince George and Northwest Applicants

Score Count Average Avrage Measure S.E. MmSq ZStd MmSq ZStd Nu Student 232 50 4.6 4.65 2.61 .26 0.7 -1 0.7 -1 1 11001 199 50 4.0 3.99 1.12 .18 1.0 0 1.0 0 2 11002 212 50 4.2 4.26 1.60 .20 1.5 2 1.4 1 3 11004 194 50 3.9 3.89 .96 .18 1.4 1 1.4 1 5 11005 199 50 4.0 3.99 1.12 .18 0.9 0 9 6 11006 207 50 4.1 4.159 1.41 .19 1.0 0 1.0 7 1007 207 50 4.1 4.15 1.41 .19 1.0 0.1 0 1.0 100100	
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	ł
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	i
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	1
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	i
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	İ
$ \begin{bmatrix} 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 1.0 & 0 & 1.0 & 0 & 8 & 11008 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.8 & -1 & 0.8 & -1 & 9 & 11009 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.1 & 0 & 1.1 & 0 & 10 & 11010 \\ 224 & 50 & 4.5 & 4.50 & 2.14 & .23 & 0.7 & -1 & 0.7 & -1 & 11 & 11011 \\ 208 & 50 & 4.2 & 4.17 & 1.44 & .19 & 0.9 & 0 & 0.9 & 0 & 12 & 11012 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.8 & -1 & 0.8 & 0 & 13 & 11013 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.2 & 0 & 1.2 & 1 & 14 & 11014 \\ 197 & 50 & 3.9 & 3.95 & 1.06 & .18 & 0.5 & -3 & 0.5 & -2 & 15 & 11015 \\ 195 & 50 & 3.9 & 3.91 & .99 & .18 & 1.1 & 0 & 1.1 & 0 & 16 & 11016 \\ 157 & 50 & 3.1 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 & 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 0.1 & 0 & 18 & 1018 \\ 188 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 18 & 1018 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 & 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & .07 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & .11 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 1.7 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & .0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & .4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3$	i
$ \begin{bmatrix} 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.8 & -1 & 0.8 & -1 & 9 11009 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.1 & 0 & 1.1 & 0 & 10 11010 \\ 224 & 50 & 4.5 & 4.50 & 2.14 & .23 & 0.7 & -1 & 0.7 & -1 & 11 11011 \\ 208 & 50 & 4.2 & 4.17 & 1.44 & .19 & 0.9 & 0 & 0.9 & 0 & 12 11012 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.8 & -1 & 0.8 & 0 & 13 11013 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.2 & 0 & 1.2 & 1 & 14 11014 \\ 197 & 50 & 3.9 & 3.95 & 1.06 & .18 & 0.5 & -3 & 0.5 & -2 & 15 11015 \\ 195 & 50 & 3.9 & 3.95 & 1.06 & .18 & 0.5 & -3 & 0.5 & -2 & 15 11015 \\ 195 & 50 & 3.9 & 3.91 & .99 & .18 & 1.1 & 0 & 1.1 & 0 & 16 11016 \\ 157 & 50 & 3.1 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 11018 \\ 188 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 11020 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 1.6 & 2 & 1.5 & 2 & 28 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & .30 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.66 & .20 & 0.7 & -1 & 0.7 & -1 & 35 11035 \\ \end{bmatrix}$	i
$ \begin{bmatrix} 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.1 & 0 & 1.1 & 0 & 10 & 11010 \\ 224 & 50 & 4.5 & 4.50 & 2.14 & .23 & 0.7 & -1 & 0.7 & -1 & 11 & 11011 \\ 208 & 50 & 4.2 & 4.17 & 1.44 & .19 & 0.9 & 0 & 0.9 & 0 & 12 & 11012 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.8 & -1 & 0.8 & 0 & 13 & 11013 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.2 & 0 & 1.2 & 1 & 14 & 11014 \\ 197 & 50 & 3.9 & 3.95 & 1.06 & .18 & 0.5 & -3 & 0.5 & -2 & 15 & 11015 \\ 195 & 50 & 3.9 & 3.91 & .99 & .18 & 1.1 & 0 & 1.1 & 0 & 16 & 11016 \\ 157 & 50 & 31 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 & 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 & 11018 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.9 & 0 & 0.9 & 0 & 19 & 11019 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 & 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & .30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	İ
$ \begin{bmatrix} 224 & 50 & 4.5 & 4.50 & 2.14 & .23 & 0.7 & -1 & 0.7 & -1 & 11 11011 \\ 208 & 50 & 4.2 & 4.17 & 1.44 & .19 & 0.9 & 0 & 0.9 & 0 & 12 11012 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.8 & -1 & 0.8 & 0 & 13 11013 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.2 & 0 & 1.2 & 1 & 14 11014 \\ 197 & 50 & 3.9 & 3.95 & 1.06 & .18 & 0.5 & -3 & 0.5 & -2 & 15 11015 \\ 195 & 50 & 3.9 & 3.91 & .99 & .18 & 1.1 & 0 & 1.1 & 0 & 16 11016 \\ 157 & 50 & 3.1 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 11018 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.9 & 0 & 0.9 & 0 & 19 11019 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 11030 \\ 175 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 11023 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 11035 \\ \end{bmatrix}$	Í
$ \begin{bmatrix} 208 & 50 & 4.2 & 4.17 \\ 188 & 50 & 3.8 & 3.77 \\ .78 & .17 & 0.8 & -1 & 0.8 & 0 & 13 & 11013 \\ 212 & 50 & 4.2 & 4.26 \\ 1.60 & .20 & 1.2 & 0 & 1.2 & 1 & 14 & 11014 \\ 197 & 50 & 3.9 & 3.95 & 1.06 & .18 & 0.5 & -3 & 0.5 & -2 & 15 & 11015 \\ 195 & 50 & 3.9 & 3.91 & .99 & .18 & 1.1 & 0 & 1.1 & 0 & 16 & 11016 \\ 157 & 50 & 3.1 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 & 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 & 11018 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.9 & 0 & 0.9 & 0 & 19 & 11019 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 & 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	
$ \begin{bmatrix} 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.8 & -1 & 0.8 & 0 & 13 & 11013 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.2 & 0 & 1.2 & 1 & 14 & 11014 \\ 197 & 50 & 3.9 & 3.95 & 1.06 & .18 & 0.5 & -3 & 0.5 & -2 & 15 & 11015 \\ 195 & 50 & 3.9 & 3.91 & .99 & .18 & 1.1 & 0 & 1.1 & 0 & 16 & 11016 \\ 157 & 50 & 3.1 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 & 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 & 11018 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.9 & 0 & 0.9 & 0 & 19 & 11019 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 & 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	ļ
$ \begin{bmatrix} 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.2 & 0 & 1.2 & 1 & 14 & 11014 \\ 197 & 50 & 3.9 & 3.95 & 1.06 & .18 & 0.5 & -3 & 0.5 & -2 & 15 & 11015 \\ 195 & 50 & 3.9 & 3.91 & .99 & .18 & 1.1 & 0 & 1.1 & 0 & 16 & 11016 \\ 157 & 50 & 3.1 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 & 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 & 11018 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.9 & 0 & 0.9 & 0 & 19 & 11019 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 & 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & 20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	
$ \begin{bmatrix} 197 & 50 & 3.9 & 3.95 \\ 195 & 50 & 3.9 & 3.91 \\ 195 & 50 & 3.1 & 3.16 \\ 157 & 50 & 3.1 & 3.16 \\ 1.57 & 50 & 4.1 & 4.09 \\ 1.30 & 1.9 & 1.5 & 2 & 1.5 & 2 \\ 1.5 & 2 & 1.7 & 11017 \\ 204 & 50 & 4.1 & 4.09 \\ 1.30 & 1.9 & 1.0 & 0 & 1.1 & 0 \\ 188 & 50 & 3.8 & 3.77 \\ 206 & 50 & 4.1 & 4.13 \\ 1.37 & 1.9 & 0.6 & -2 & 0.7 & -1 \\ 206 & 50 & 4.1 & 4.13 \\ 1.37 & 1.9 & 0.6 & -2 & 0.7 & -1 \\ 207 & 50 & 4.1 & 4.15 \\ 1.41 & 1.9 & 0.9 & 0 & 0.9 & 0 \\ 180 & 50 & 3.6 & 3.62 \\ 2.55 & 1.7 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 \\ 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 \\ 144 & 50 & 3.7 & 3.68 \\ .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 \\ 163 & 50 & 3.3 & 3.28 \\ .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 \\ 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 \\ .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	
$ \begin{bmatrix} 195 & 50 & 3.9 & 3.91 & .99 & .18 & 1.1 & 0 & 1.1 & 0 & 16 & 11016 \\ 157 & 50 & 3.1 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 & 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 & 11018 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.9 & 0 & 0.9 & 0 & 19 & 11019 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 & 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11023 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{array}$	
$ \begin{bmatrix} 157 & 50 & 3.1 & 3.16 &03 & .15 & 1.5 & 2 & 1.5 & 2 & 17 & 11017 \\ 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 & 11018 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.9 & 0 & 0.9 & 0 & 19 & 11019 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 & 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	
$ \begin{bmatrix} 204 & 50 & 4.1 & 4.09 & 1.30 & .19 & 1.0 & 0 & 1.1 & 0 & 18 11018 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 0.9 & 0 & 0.9 & 0 & 19 11019 \\ 206 & 50 & 4.1 & 4.13 & 1.37 & .19 & 0.6 & -2 & 0.7 & -1 & 20 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 11035 \\ \end{bmatrix}$	[
$ \begin{bmatrix} 188 & 50 & 3.8 & 3.77 \\ 206 & 50 & 4.1 & 4.13 \\ 1.37 & .19 & 0.6 & -2 & 0.7 & -1 \\ 20 & 1020 \\ 180 & 50 & 3.6 & 3.62 \\ 207 & 50 & 4.1 & 4.15 \\ 1.41 & .19 & 0.9 & 0 & 0.9 & 0 \\ 22 & 11021 \\ 207 & 50 & 4.1 & 4.15 \\ 1.41 & .19 & 0.9 & 0 & 0.9 & 0 \\ 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 \\ 183 & 50 & 3.7 & 3.68 \\ .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 \\ 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 \\ .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 \\ 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 \\ .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	
$ \begin{bmatrix} 205 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & .1 & 0 & 21 & 11020 \\ 180 & 50 & 3.6 & 3.62 & .55 & .17 & 1.1 & 0 & 1.1 & 0 & 21 & 11021 \\ 207 & 50 & 4.1 & 4.15 & 1.41 & .19 & 0.9 & 0 & 0.9 & 0 & 22 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	
$ \begin{bmatrix} 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 &5 & 0 & 2.2 & 11022 \\ 144 & 50 & 2.9 & 2.90 &32 & .15 & 1.2 & 0 & 1.2 & 0 & 23 & 11023 \\ 183 & 50 & 3.7 & 3.68 & .63 & .17 & 0.8 & -1 & 0.8 & 0 & 24 & 11024 \\ 203 & 50 & 4.1 & 4.07 & 1.26 & .19 & 0.5 & -2 & 0.5 & -2 & 25 & 11025 \\ 163 & 50 & 3.3 & 3.28 & .11 & .15 & 0.7 & -1 & 0.7 & -1 & 26 & 11026 \\ 212 & 50 & 4.2 & 4.26 & 1.60 & .20 & 1.7 & 2 & 1.6 & 2 & 27 & 11027 \\ 188 & 50 & 3.8 & 3.77 & .78 & .17 & 1.6 & 2 & 1.5 & 2 & 28 & 11028 \\ 202 & 50 & 4.0 & 4.05 & 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix}$	
142502.52.51.521.201.202.511025183503.73.68.63.17 0.8 -1 0.8 02411024203504.14.071.26.19 0.5 -2 0.5 -22511025163503.33.28.11.15 0.7 -1 0.7 -12611026212504.24.261.60.20 1.7 2 1.6 22711027188503.8 3.77 .78.17 1.6 2 1.5 22811028202504.0 4.05 1.23 .19 1.1 0 1.1 0291102918150 3.6 3.64 .57.17 0.7 -1 0.7 -1 301103017550 3.5 3.52 .41.16 0.9 0 1.0 0 311103119250 3.8 3.85 .90.18 0.5 -2 0.5 -2 32 1103217950 3.6 3.60 52 .17 1.5 1 1.4 1 33 1103321650 4.3 4.34 1.76 .21 0.6 -2 34 1103421450 4.3 4.30 1.68 .20 0.7 -1 0.7 -1 35 11035 </td <td></td>	
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	
$ \begin{bmatrix} 163 & 50 & 3.3 & 3.28 \\ 163 & 50 & 4.2 & 4.26 \\ 212 & 50 & 4.2 & 4.26 \\ 212 & 50 & 4.2 & 4.26 \\ 212 & 50 & 4.2 & 4.26 \\ 212 & 50 & 4.2 & 4.26 \\ 212 & 50 & 4.2 & 4.26 \\ 212 & 50 & 4.2 & 4.26 \\ 202 & 50 & 4.0 & 4.05 \\ 202 & 50 & 4.0 & 4.05 \\ 202 & 50 & 4.0 & 4.05 \\ 1.23 & .19 & 1.1 & 0 & 1.1 & 0 & 29 & 11029 \\ 181 & 50 & 3.6 & 3.64 & .57 & .17 & 0.7 & -1 & 0.7 & -1 & 30 & 11030 \\ 175 & 50 & 3.5 & 3.52 & .41 & .16 & 0.9 & 0 & 1.0 & 0 & 31 & 11031 \\ 192 & 50 & 3.8 & 3.85 & .90 & .18 & 0.5 & -2 & 0.5 & -2 & 32 & 11032 \\ 179 & 50 & 3.6 & 3.60 & .52 & .17 & 1.5 & 1 & 1.4 & 1 & 33 & 11033 \\ 216 & 50 & 4.3 & 4.34 & 1.76 & .21 & 0.6 & -2 & 0.6 & -2 & 34 & 11034 \\ 214 & 50 & 4.3 & 4.30 & 1.68 & .20 & 0.7 & -1 & 0.7 & -1 & 35 & 11035 \\ \end{bmatrix} $	
212 50 4.2 4.26 1.60 .20 1.7 2 1.6 2 27 11027 188 50 3.8 3.77 .78 .17 1.6 2 1.5 2 28 11027 188 50 3.8 3.77 .78 .17 1.6 2 1.5 2 28 11028 202 50 4.0 4.05 1.23 .19 1.1 0 1.1 0 29 11029 181 50 3.6 3.64 .57 .17 0.7 -1 0.7 -1 30 11030 175 50 3.5 3.52 .41 .16 0.9 0 1.0 0 31 11031 192 50 3.8 3.85 .90 .18 0.5 -2 0.5 -2 32 11032 179 50 3.6 3.60 .52 .17 1.5 1 1.4 1 33 11033 216 50 4.3 4	
188 50 3.8 3.77 .78 .17 1.6 2 1.5 2 28 11028 202 50 4.0 4.05 1.23 .19 1.1 0 1.1 0 29 11028 202 50 4.0 4.05 1.23 .19 1.1 0 1.1 0 29 11029 181 50 3.6 3.64 .57 .17 0.7 -1 0.07 -1 30 11030 175 50 3.5 3.52 .41 .16 0.9 0 1.0 0 31 11031 192 50 3.8 3.85 .90 .18 0.5 -2 0.5 -2 32 11032 179 50 3.6 3.60 .52 .17 1.5 1 1.4 1 33 11033 216 50 4.3 4.30 1.68 .20 0.7 -1 0.7 -1 35 11034 214 50 4.3 <	
202 50 4.0 4.05 1.23 .19 1.1 0 1.1 0 29 11029 181 50 3.6 3.64 .57 .17 0.7 -1 0.7 -1 30 11030 175 50 3.5 3.52 .41 .16 0.9 0 1.0 0 31 11031 192 50 3.8 3.85 .90 .18 0.5 -2 0.5 -2 32 11032 179 50 3.6 3.60 .52 .17 1.5 1 1.4 1 33 11033 216 50 4.3 4.34 1.76 .21 0.6 -2 0.6 -2 34 11034 214 50 4.3 4.30 1.68 .20 0.7 -1 0.7 -1 35 11035	i
181 50 3.6 3.64 .57 .17 0.7 -1 0.7 -1 30 11030 175 50 3.5 3.52 .41 .16 0.9 0 1.0 0 31 11031 192 50 3.8 3.85 .90 .18 0.5 -2 0.5 -2 32 11032 179 50 3.6 3.60 .52 .17 1.5 1 1.4 1 33 11033 216 50 4.3 4.34 1.76 .21 0.6 -2 0.6 -2 34 11034 214 50 4.3 4.30 1.68 .20 0.7 -1 0.7 -1 35 11035	i
175 50 3.5 3.52 .41 .16 0.9 0 1.0 0 31 11031 192 50 3.8 3.85 .90 .18 0.5 -2 0.5 -2 32 11032 179 50 3.6 3.60 .52 .17 1.5 1 1.4 1 33 11033 216 50 4.3 4.34 1.76 .21 0.6 -2 0.6 -2 34 11034 214 50 4.3 4.30 1.68 .20 0.7 -1 0.7 -1 35 11035	i
192 50 3.8 3.85 .90 .18 0.5 -2 0.5 -2 32 11032 179 50 3.6 3.60 .52 .17 1.5 1 1.4 1 33 11033 216 50 4.3 4.34 1.76 .21 0.6 -2 0.6 -2 34 11034 214 50 4.3 4.30 1.68 .20 0.7 -1 0.7 -1 35 11035	Í
179 50 3.6 3.60 .52 .17 1.5 1 1.4 1 33 11033 216 50 4.3 4.34 1.76 .21 0.6 -2 0.6 -2 34 11034 214 50 4.3 4.30 1.68 .20 0.7 -1 0.7 -1 35 11035	i
216 50 4.3 4.34 1.76 .21 0.6 -2 0.6 -2 34 11034 214 50 4.3 4.30 1.68 .20 0.7 -1 0.7 -1 35 11035	İ
	1
175 50 3.5 3.52 .41 .16 0.6 -2 0.6 -2 36 11036	
193 50 3.9 3.87 .93 .18 0.8 -1 0.8 -1 37 11037	
196 50 3.9 3.93 1.03 .18 1.0 0 1.0 0 38 21038	
224 50 4.5 4.50 2.14 .23 1.1 0 1.1 0 39 21039	
$\begin{bmatrix} 199 & 50 & 4.0 & 3.99 \end{bmatrix} 1.12 & 18 \end{bmatrix} 1.0 & 0 & 1.0 & 0 \end{bmatrix} 40 21040$	
104 - 20 - 3 0 3 00 05 - 10 05 - 0 0 3 - 1 14 01044 120 - 20 - 4·0 - 2·2/ 1·02 - 12 1·1 0 - 1·7 0 43 51043	
157 40 3.9 3.94 1.03 .20 0.8 -1 0.7 -1 50 22050	l
183 40 4.6 4.59 2.40 .27 1.2 0 1.4 1 51 22051	i
124 32 3.9 3.90 .97 .22 0.6 -1 0.7 -1 52 22052	i
168 40 4.2 4.21 1.52 .22 0.7 -1 0.7 -1 53 22053	i
177 40 4.4 4.44 2.00 .24 0.9 0 0.9 0 54 22054	ĺ
168 40 4.2 4.21 1.52 .22 1.4 1 1.4 1 55 22055	l
177 40 4.4 4.44 2.00 .24 0.5 -2 0.5 -2 56 22056	Í
165 40 4.1 4.14 1.38 .21 1.2 0 1.2 0 57 22057	Í
152 40 3.8 3.81 .84 .20 1.4 1 1.4 1 58 22058	
154 40 3.9 3.86 .91 .20 0.9 0 0.9 0 59 22059	- 1
174 40 4.3 4.36 1.83 .23 1.0 0 0.9 0 60 22060	ļ
167 40 4.2 4.19 1.47 .22 0.9 0 0.9 0 61 22061	ļ
171 40 4.3 4.29 1.67 .23 1.0 0 1.0 0 62 22062	ļ
1 184 40 4.6 4.61 2.48 .28 0.8 0 0.9 0 64 22064	

ļ	Obsvđ	Obsvđ	Obsvd	Fair		Model	Infi	lt	Out	Eit			ł
	Score	Count	Average	Avrge	Measure	S.E.	MnSq	zStđ	MnSq	zstd	Nu	Student	- 1
-						Conti	inued						
1	132	32	4.1	4.15	1.39	.24	1.3	1	1.5	1	65	22065	I
Í	147	40	3.7	3.69	.65	.19	1.6	2	1.7	2	66	22066	i
Ì	164	40	4.1	4.11	1.33	.21	1.1	0	1.1	0	67	22067	Í
Í	160	40	4.0	4.01	1.16	.21	0.9	0	0.9	0	68	22068	Í
ł	154	40	3.9	3.86	.91	.20	1.1	0	1.1	0	69	22069	Ì
Ì	156	40	3.9	3.91	.99	.20	0.8	0	0.8	0	70	22070	Í
ī			oh										
1	Dvadu	ODSVQ	ODSVd	Fair		Model	Inti		Outs	tit		·	
I	Score	Count	Average	Avrge	Measure	S.E.	MnSq	zstd	MnSq	zstd	Nu	Student	
1	186.2	46	.8 4.0	4.00	1.19	.20	1.0	-0.2	1.0	-0.1	Mea	an (Count: 70)	
	22.7	5	.1 0.3	0.32	.54	.03	0.3	1.5	0.3	1.4	s. 1	.	
-		RMSE Fi R	(Model) xed (all andom (n	.20 i same) ormal)	Adj S.D. chi-squa chi-squa	.51 re: 549 re: 68.	Separ 9.9 d .0 d.	ation .f.: 6	2.57 59 si 8 sig	7 Reli gnifica nificar	abi ance	Lity .87 : .00 .48	

Cat	Step	Ежр.	Resd	StRes	Nu	Stude	N	Ra	Nu	It	-
1	1	4.0	-3.0	-3	3	11003	5	05	4	4	Ī
1	1	3.6	-2.6	-3	5	11005	5	05	4	4	İ
1	1	3.7	-2.7	-3	49	21049	4	04	1	1	İ
2	2	4.2	-2.2	-3 İ	27	11027	1	01	6	6	Í
3	3	4.7	-1.7	-3 j	39	21039	1	01	7	7	İ
1	1	3.6	-2.6	-3	50	22050	1	01	4	4	İ
3	3	4.8	-1.8	-3	51	22051	1	01	5	5	İ
3	3	4.7	-1.7	-3 İ	51	22051	1	01	7	7	İ
1	1	4.0	-3.0	-3	3	11003	3	03	4	4	Í
2	2	4.2	-2.2	-3	29	11029	3	03	9	9	İ
Cat	Step	Ехр.	Resđ	StRes	Nu	Stude	N	Ra	Nu	It	

Appendix F: Unexpected Responses Report for Prince George and Northwest Applicants