

An Empirical Assessment of the Proposed Trent University Course Faculty Evaluation Questionnaire

*Diana Peglar, Kirsten Marchand,
Rachel Grills, Kelly Roberts*

Abstract: The present study examined the psychometric characteristics of the proposed CFEQ – an interdepartmental measure of teaching effectiveness. Fifty three undergraduate Psychology students completed the 60-item CFEQ, the current Trent University Psychology Teaching Evaluation Questionnaire, and the WICE. Results of the present investigation revealed that 55 of the 56 proposed items for the CFEQ had high internal consistency. The scale shows high reliability, more than sufficient for clinical or diagnostic purposes, and when compared to three criterion measures of teaching effectiveness, substantial correlations were demonstrated for the CFEQ, suggesting very strong construct validity. In sum, the proposed CFEQ appears to be a very reliable and valid assessment measure for tapping the pertinent categories of effective teaching. Therefore, this innovative tool has excellent potential for future standardization across the various disciplines within Trent University.

Presently, many methods are in use for evaluating the effectiveness of teachers at a post-secondary level. Possible methods include teacher examinations, classroom observations, interviews and portfolio assessment (Flowers & Hancock, 2003). However, student evaluations are still considered to be effective, reliable and valid measures for examining the quality of instruction given at the post-secondary level (Vernooij-Dassen, 2000, as cited in Turhan, Yaris, & Nural, 2005; Wilkes & Bligh, 1999, as cited in Krantz-Girod, Bonvin, Lanares, Cuénot, & Feihl, 2004). Feedback from student evaluations is beneficial to both the faculty and the students because it provides a source of information which instructors can use to improve the quality of instruction (Wachatel, 1998, as cited in Turhan *et al.*, 2005), increase student motivation (Wlodkowski, 1999, as cited in Turhan *et al.*, 2005), and provide information on which to make tenure, promotions or resource allocation decisions (Lancaster *et al.*, 1979, Smith & Paulen, 1984, Elzubeir & Rizk, 2002, all as cited in Krantz-Girod *et al.*, 2004).

Davey and Neil (1991, as cited in Flowers & Hancock, 2003) suggest that teaching is a dynamic interplay of complex sets of personal attributes, knowledge and abilities. This makes the measurement of the construct “effective teaching” a difficult task. The Interstate New Teacher Assessment and Support Consortium (INTASC, 1995, as cited in Flowers & Hancock, 2003) proposed that an effective teacher should be able to: (1) adopt flexible teaching strategies that benefit different learning styles, (2) fully understand the subject matter of the course in order to relate it to their students, (3) use a wide variety of instructional strategies in the classroom, (4) create proper tools of evaluation to assess student learning, and (5) continually participate in curriculum and instructional evaluation and development. Other factors commonly identified as characteristics of effective teaching include clarity, stimulation, course difficulty, organization, interaction with the students, and overall rating (Marsh, 1982, as cited in Krantz-Girod *et al.*, 2004).

The current measures of teaching effectiveness in use at Trent University all consider similar traits to the characteristics mentioned above, though there is no consensus within the

departments as to the best method to evaluate these central traits. In 1974, Trent University proposed a standardized questionnaire to be used for all courses in every university department. The proposed 20-item scale was administered to nine courses from different departments within the university and was found to be both reliable and valid (Reker, Jury, & Wadland, 1976). However, this standardized measure was never put into use, and to this day, each department within Trent has its own rules and questionnaires for evaluating teaching effectiveness.

The purpose of the present study is to conduct a revision to the evaluation forms presently used by Trent University to decipher the most suitable and functionally significant questions that represent an appropriate evaluation survey. Our ultimate goal is that this will lead to the development of a standardized course faculty evaluation to be used across all departments within Trent University. This would provide the university with a general tool for assessing any given instructor’s teaching effectiveness and assist in personnel decision-making. The use of a standardized course faculty evaluation would also allow for interdepartmental comparisons in teaching effectiveness.

Method

Participants

Participants included a group of 53 undergraduate Psychology students from a third year Statistics course (PSYC 315) and a third year Psychometrics course (PSYC 332) at Trent University. All participants were volunteers. The sample included both male and female students. No socio-demographic data was collected.

Measures

There were three standardized measures administered in total. Each participant was given a package including the 60-item Course Faculty Evaluation Questionnaire (CFEQ), the current 19-item Trent University Psychology Teaching Evaluation Questionnaire (TEQ), and the 16-item University of Western Ontario Instructor/ Course Evaluation (WICE). In accordance with the University’s ethical regulations the subjects were also provided with a Consent Form and Feedback Sheet.

First in the package was the target CFEQ, created by the researchers as a representation of all the valuable traits comprised in “effective teaching”. It is currently under investigation in this study for reliability and validity as a standardized measure that could be used across disciplines within Trent University. Encompassed in this questionnaire were four items that were hypothesized to be an accurate summary of the overall effectiveness of the instructor evaluated. Those four items included the following questions: Item 57 (“How would you rate the instructor’s overall teaching effectiveness?”); Item 58 (“Would you take another course taught by this instructor?”); Item 59 (“Would you recommend this course to other students?”); and Item 60 (“Would you recommend this instructor teaches this course again next year?”). The researchers chose (a priori to statistical comparisons) to treat these four items separately from the remaining 56-items on the scale and use these variables as additional criterion measures. These items were compared by a Pearson’s *r* correlation coefficient on the STATISTICA software program to uncover any significant inter-correlations. The inter-correlations were all significant at an alpha level of .01 or higher; however most were still significant at a much more conservative alpha level. Subsequently, this formed the logical decision base for combining these four items into a new variable of the Overall Instructor Rating (4ITEMTOT). This newly created single variable was used as the third criterion measure for assessing the convergent evidence of construct validity.

Second in the package was the TEQ. This was previously shown to be both reliable and valid for standardized psychology evaluation (Reker, Jury, & Wadland, 1976). The reliability estimate for the TEQ was .81 significant at a conservative alpha level of .001 (Reker, Jury, & Wadland, 1976). The validity ($r = .61, p < .001$) was established through concurrent evidence of construct validity by comparing the TEQ to the University of Waterloo questionnaire (Reker, Jury, & Wadland, 1976). For the purposes of this research, the TEQ was used as an external criterion measure for comparison to the target scale (CFEQ) for evidence of convergent validity.

Last in the package was the WICE, which was also used as a criterion measure for validating the target scale. The WICE was assumed to be reliable and valid since it is presently used as a standardized measure of course and instructor effectiveness at this particular educational institution. Unfortunately, for this study we did not have access to any specific evidence to verify the reliability and validity of the scale.

Procedure

The typical performance CFEQ was created using a set of characteristics proposed by the Chair of the Psychology Department, at Trent University (see Appendix for the complete list). After accounting for these mandatory traits and examining the current evaluations used by each department at Trent University, we assembled a list of approximately 4 stems that represented each of the fundamental categories. These stems were then compiled into a randomized questionnaire following a 5-point Likert scale format. The possible responses included ratings of Poor (1), Fair (2), Good (3), Very Good (4) and Excellent (5). We included more positive than negative response items since previous research has found that students are likely to consider even the poorest professors at Trent as average (Reker, Jury, & Wadland, 1976). This is due to the fact that professors are required to surpass a minimum standard of excellence in order to be employed with the

University.

Demographic information such as age, sex, program, major and courses completed to date was excluded from the construction of the proposed CFEQ. The reasoning for this was that prior research found their effects on faculty ratings to be insignificant (Reker, Jury, & Wadland, 1976). As a result, any such information present on either criterion measure (TEQ and WICE) was not included for the data analyses.

Upon approval from the Psychology Ethics Committee, participants were recruited from PSYC 315 and PSYC 332 classes. Questionnaire packages were completed directly after the lecture for PSYC 315 volunteers, whereas for PSYC 332 students, time was allotted during lecture to complete the surveys.

Participants were instructed to complete the three teaching effectiveness questionnaires (CFEQ, TEQ and WICE) and return them to one of the four researchers. Participants were not asked to provide their names or student numbers to insure their anonymity during the research process. The subjects were also given a Consent Form prior to their completion of the questionnaire package, and handed a Feedback Sheet after submission. When the questionnaires were submitted, they were assigned an arbitrary identification number, which was used to code the data.

Using the STATISTICA program an item analysis was performed using the Pearson’s Product Moment Correlation Coefficient, followed by an analysis of reliability using Cronbach’s alpha and Split-Half Reliability. Finally, the construct validity of the CFEQ was determined by a convergent analysis correlating the Target Scale Total to three criterion measures: the TEQ total, WICE total, and Overall Rating total derived from the last four items (57-60) of the target scale. Results of these analyses are presented below.

Results

Item Analysis

The item total correlation method revealed that items 1-52 and 54-56 have high correlations with the total test score, and thus are items that will be retained. The correlation coefficients for each of these items with the total test score can be found in Appendix A. With the exception of a few items, most are significant with 99.9% confidence. Items 30, 34, 48, and 50, although still significant, differ from the remaining items, as they are significant at either 99 or 95% confidence levels. Respectively, the correlations for these items are as follows: $r = .45, p < .01$; $r = .31, p < .05$; $r = .40, p < .01$ and $r = .40, p < .01$.

Item 53 (i.e. “How would you rate the instructor’s ability in incorporating personal research into the lectures/seminars?”) deserves special mention as it is the only item on the questionnaire, which has an insignificant correlation coefficient. The coefficient for item 53 is $r = .17, p > .05$. As a result, item 53 is considered a “bad” item and will be discarded from further analyses. It will not appear on the final questionnaire.

Reliability

The reliability estimates of the CFEQ are presented in Table 1. Cronbach’s alpha reliability estimate for the CFEQ was calculated using the 55 “good” items (Items 1-52 and 54-56). Cronbach’s alpha

Table 1
Reliability of the Course Faculty Evaluation Questionnaire (N=53)

| | |
|---|-----|
| Course Faculty Evaluation Questionnaire | |
| Cronbach’s Alpha | .97 |
| Split-half Reliability | .98 |

Table 2

Convergent Evidence of Construct Validity through Three External Criterion Measures (N=53)

| Criterion Measures | Course Faculty Evaluation |
|---|---------------------------|
| Trent Psychology Department Teaching Evaluation Questionnaire (TEQ) | .92 *** |
| University of Western Ontario Course and Instructor Evaluation (WICE) | .78 *** |
| Overall Rating (4ITEMTOT) | .74 *** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

determines how well a set of items measures a unidimensional, latent construct; thus this coefficient of reliability was used to test for the questionnaire’s internal consistency. The Cronbach’s alpha attained for the CFEQ was .97. This exceeds the significance requirements for research and diagnostic purposes (.70 and .90 respectively). This demonstrates a high inter-item correlation, illustrating that the items are measuring the same underlying construct. As a result of the questionnaire’s high reliability, deleting and/or adding items will not cause a considerable change in the reliability, and consequently all 55 items will be retained for the CFEQ.

The split-half reliability of the CFEQ (as adjusted by Spearman-Brown) was .98. The items were divided using the odd-even split technique. There was no significant difference between the two reliability estimates, although Cronbach’s alpha is generally a more conservative measure of reliability as it splits the test in all possible ways. This justifies the minor difference between the two values.

Validity

Since the CFEQ revealed item to item consistency, the scale was subsequently assessed for the most sophisticated type of validity. Construct validity has been conceptualized as the most constructive and functionally useful representation of validity (Cohen & Swerdlik, 2005). Therefore, three external criterion measures that tapped an identical (or similar) representation of teaching effectiveness were utilized to evaluate the CFEQ. Consequently, it was possible to gather convergent evidence of construct validity.

The results of the Pearson’s Product Moment Correlation Coefficient between the proposed CFEQ and the three external criterion measures revealed significantly high validity correlation coefficients. These correlations are presented in Table 2. The correlation between the current TEQ and the target CFEQ was shown to be the highest, $r = .92$, $p < .001$; however all three correlations were significant at the alpha level of .001. Thus, convergent evidence of empirical criterion validity was successfully shown.

The coefficient of determination was used to calculate the percent of shared variance in teaching effectiveness for the three external criterion measures. The validity coefficients are consistently significant, whereby the Overall Rating accounted for 55%, the TEQ accounted for 85% and WICE accounted for 61% of the variance. In sum, the TEQ was the greatest representation for empirical construct validation of the CFEQ.

Discussion

From the statistical analyses presented above we can see that the proposed CFEQ has been revealed to be a reliable and valid tool for the assessment of teaching effectiveness. Aside from Item 53, all other items on the scale were found to be “good”, and have

high internal consistency. In order to truly determine if the CFEQ is reliable and valid for standardization across all departments at Trent University, a second administration over time, across disciplines and years of study is necessary to ensure stability. Due to the time available for the current research, this in

depth procedure was not possible but should be considered for future investigations.

For the simplicity of our research purposes, we chose to use a paper and pencil format for the CFEQ. Although some departments in the university currently use a web-based response for teaching effectiveness evaluations, we advise against this type of format for the CFEQ. Prior research has determined that response rates are lower on web-based questionnaires than when students fill out a paper and pencil form in class (Turhan *et al.*, 2005). Some classes offer a bonus mark for completing online surveys; however more research would be needed to determine whether this actually increases the number of students who complete the questionnaires. Allowing instructors to offer bonus marks raises ethical concerns, as it is likely that those instructors would be ranked more favourably among students.

Based on the results from the item analysis, it appears that there is a very strong correlation for all the items (excluding Item 53) with the total test score. This implies that the items developed for the CFEQ are good representations of the effectiveness of teaching for the particular courses and instructors sampled.

To address any problems caused by ambiguous items, or those that do not apply to a particular course and/or instructor, it may be useful to include a sixth response category of “not applicable” (i.e. N/A) to the original 5-point Likert scale. Future samples would be asked to use such a category only when necessary to avoid any central tendency response biases. The benefit of adding a N/A category is that instructors will not be unfairly rated due to the nature of their curriculum. With the addition of the N/A category, students are able to exclude a question when it does not apply to that particular course.

Previous research conducted by Reker *et al.* (1976) also found that the exclusion of a N/A category creates response patterns where students would either leave the item blank, or write N/A next to the statement. These researchers suggested that the inclusion of the N/A category would eliminate this type of response pattern. They believed it would be beneficial to include the following instructions: “If a particular question is not applicable to the course and/or the instructor, check the NA response.” These suggestions are consistent with those made above by the current researchers.

Due to time constraint, test-retest was not practical in calculating the reliability for the CFEQ. However, time permitted it would have been ideal to complete a second administration of the CFEQ in order to obtain the test-retest reliability and determine the stability of students response patterns over time. Stability estimates of reliability are important to ensure that the questionnaire is tapping the same traits of instructors’ teaching effectiveness year after year. This stability estimate can also be obtained by administering the CFEQ for any given instructor twice within the same semester. The amount of time allotted between the two assessments is

critical. Consequently, if a second administration of the CFEQ was completed, a minimum four weeks between trials would be optimal for data collection.

Again due to time constraints, delayed parallel forms could not be constructed or administered. In order to do so, a factor analysis is necessary in order to decipher the appropriate items to be used for each version of the questionnaire. For future consideration two versions of the CFEQ could be developed based on the results of our item-analysis. The parallel forms could each include 20 to 25-items since the original CFEQ contains 55 "good" items. The use of two shorter scales would be ideal, as students and instructors cannot afford to spend the time necessary to complete a 60-item scale.

The parallel forms could be created via a ranked list from highest to lowest correlation coefficients and then isolating each related pair of items and randomly separating them into Form A and Form B. Consequently, each form would have very similar item total correlations, reliability, and validity. Having parallel forms that measure the same construct is beneficial if standardized across Trent University, as all professors would be evaluated equally under the same principles resulting in a more fair assessment procedure.

A valid measure of "teaching effectiveness" should be highly correlated with other criterion measures that also assess the construct. The results reveal that this is precisely the case, implying that the validity is highly acceptable. Additionally, the last four items of the proposed CFEQ also empirically validated the construct representation. These four items could also be used to validate the homogeneity of the questionnaire in its measurement of a single concept; however they were not manipulated in such a manner for the present study.

Convergent validity was the only analysis possible to employ in the present study given the limited resources (i.e. volunteer commitment) and time. A problem with this type of validity check is that it raises certain concerns about the originality of the items. Basically, it begs to question whether the new evaluation simply duplicates various other existing measures of the same construct. To adjust for this type of limitation, and maximize the validity checks of the CFEQ, predictive, and discriminant validity should also be assessed.

Prospectively, to gather divergent evidence of construct validity a comparison of our CFEQ to unrelated variables would be useful. Additionally, investigating the response phenomenon of Social Desirability is a pertinent aspect for future research. The students may have decided to respond with the intention of pleasing their professor in hopes of leniency with regards to grading or to receive special consideration. However this response style may have been minimized, since our questionnaire specifically stated that the surveys were anonymous and that the professors evaluated would not have access to the forms until after the final course marks were allocated.

Likewise, a predictive validity check could have been completed using the records of previous Symons Award winners as the predictor variable, and the ratings of these winners on the CFEQ as the criterion measure. It could be hypothesized that winners of this award would receive higher overall scores on the CFEQ which would produce additional evidence for related construct validity. If this was not the case it could be plausible that past winners of the Symons Award may have been confounded by extraneous factors such as the grades awarded to the students enrolled in their classes. This is problematic since these professors may

be buying favourable ratings through their grading schema. To control for different grading stringency (i.e. Easy, Normal or Hard Marker), a regression analysis could be completed with grades given as the covariate. This would in turn partial out the effects of grades on teaching effectiveness and provide an adjusted value for the ratings. Previous Trent research found that a higher obtained grade typically ensured a higher ranking on teaching effectiveness, independent of other factors (Reker *et al.*, 1976). Therefore, it would be beneficial to include questions such as "Expected grade in the course?" and "Grade obtained to this point?"

Reker *et al.* (1976) also found that high attendance at lecture resulted in more positive ratings of both the instructor and the assigned readings. For future consideration, it would be valuable to include questions such as "Percentage of lectures attended?" "Percentage of tutorials attended?", and "Percentage of labs attended?" in order to fully understand the response patterns of students.

In a study by Krantz-Girod *et al.* (2004) it was found that when open-ended questions (e.g. "Any additional comments?") appear on evaluations, the percentage of negative comments is much higher relative to the percentage of positive comments. The researchers suggest that students are likely to take action when unsatisfied since it is their right to effective instruction. Although one open-ended question was present on the CFEQ, no analyses were completed on any comments made. In the future, it would be advantageous to compare these comments to those on the TEQ, and analyse the qualitative relationships in comparison to quantitative data.

Although teaching effectiveness evaluations are highly valued as a way to analyze post secondary instructor's performance in the classroom, do these questionnaires make a difference? All instructors have access to these forms in order to make use of the comments to see where any problems may lie. But do they use this to proactively take charge and change their instruction to improve the learning that takes place in their classrooms?

Krantz-Girod *et al.* (2004), Cohen (1980, as cited in Krantz-Girod *et al.*, 2004), Ory (2000, as cited in Krantz-Girod *et al.*, 2004), and Kulik (2001, as cited in Krantz-Girod *et al.*, 2004) all found that repeated evaluations of teaching effectiveness are not enough of a catalyst to improve the quality of instruction. However, Diamond (2004) found that when used earlier in the term (i.e. before the midterm), student evaluations increased the instructor's awareness of areas necessary for improvement and resulted in amendments to these matters. In order for the CFEQ to be successful, Trent University may want to standardize the questionnaire and administration procedures. This could maximize the utility of students' opinions on teaching effectiveness and allow for an objective nomination for instructor awards (i.e. Symons Award) across departmental divisions.

Results of the present investigation revealed that 55 of the 56 proposed items for the CFEQ had high internal consistency. Subsequent analyses showed that the reliability for those remaining items was more than sufficient for clinical or diagnostic purposes. When compared to three other criterion measures of teaching effectiveness, substantial correlations were demonstrated for the CFEQ, suggesting very strong construct validity. In sum, the proposed CFEQ appears to be a very reliable and valid assessment measure for tapping the pertinent categories of effective teaching. Therefore, this innovative tool has excellent potential for future standardization across various disciplines within Trent University.

References

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment* (6th ed.). New York, NY: McGraw-Hill Companies Inc.

Diamond, M. R. (2004). The usefulness of structured mid-term feedback as a catalyst for change in higher education classes. *Active Learning in Higher Education*, 5, 217-231.

Flowers, C. P., & Hancock, R. D. (2003). An interview protocol and scoring rubric for evaluating teacher performance. *Assessment in Education*, 10, 161-168.

Krantz-Girod, C., Bonvin, R., Lanares, J., Cuénot, S., Feihl, F., Bosman, F., & Waeber, B. (2004). Stability of repeated student evaluations of teaching in the second preclinical year of a medical curriculum. *Assessment & Evaluation in Higher Education*, 29, 123-133.

Reker, G. T., Jury, J. W., & Wadland, J. H. (1976). *An empirical assessment of the Trent Course-Faculty Evaluation Questionnaire*. Trent University, Department of Psychology.

Turhan, K., Yaris, F., & Nural, E. (2005). Does instructor evaluation by students using a web-based questionnaire impact instructor performance? *Advances in Health Sciences Education*, 10, 5-13.

Appendix A

Item-Analysis of Course Faculty Evaluation Scores using the Item-Total Correlation Method

| Item Number | Item-Total Correlation |
|---|------------------------|
| 1. Coming to lectures and seminars well prepared? | .51 *** |
| 2. Answering questions posed by students during class? | .46 *** |
| 3. Having organized lecture material? | .67 *** |
| 4. Having lectures flow systematically according to topics? | .69 *** |
| 5. Using appropriate audio aids if necessary? | .47 *** |
| 6. Speaking with appropriate pronunciation and grammar? | .50 *** |
| 7. Using voice intonation to engage students in course material? | .66 *** |
| 8. Using clear, helpful, visual and auditory tools? | .75 *** |
| 9. Beginning and ending class at regular scheduled times? | .47 *** |
| 10. Giving ample notice if class is going to be cancelled? | .52 *** |
| 11. Outlining course requirements clearly? | .72 *** |
| 12. Maintaining accord between course objectives and what is taught? | .69 *** |
| 13. Clearly organizing the syllabus and following the outlined schedule? | .56 *** |
| 14. In directly stating expectations for students' achievement? | .65 *** |
| 15. Making workload appropriate to level of study? | .53 *** |
| 16. Making independent workload reasonable? | .51 *** |
| 17. Setting the number of assignments to reflect course curriculum? | .48 *** |
| 18. Setting an appropriate level of difficulty (projects, seminars, exams)? | .56 *** |
| 19. Allotting appropriate time for comprehension of difficult material? | .79 *** |
| 20. Explaining and emphasizing key concepts? | .76 *** |
| 21. Making the breakdown of course evaluations achievable? | .55 *** |
| 22. Time allotted for term tests/exams suitable? | .60 *** |
| 23. Ensuring consistency between marking scheme and achieved grade? | .63 *** |
| 24. Making course grades fair and accurate? | .66 *** |
| 25. Having methods of evaluation reflect key subject matter? | .68 *** |
| 26. Providing fair evaluation of students overall grasp on course material? | .72 *** |
| 27. Weighting test questions appropriately to work required for the answer? | .60 *** |
| 28. Maintaining grading uniformity (especially with multiple markers)? | .56 *** |
| 29. Giving feedback for assignments in a timely manner? | .63 *** |
| 30. Having grades available before drop date without academic penalty? | .45 ** |
| 31. Returning major assignments before final exam? | .59 *** |
| 32. Promptly marking to provide students with their current standing? | .69 *** |
| 33. Providing constructive criticism for areas of improvement? | .66 *** |
| 34. Making exams available to review errors? | .31 * |
| 35. Encouraging students to form judgments and opinions about topics? | .65 *** |
| 36. Promoting critical thinking (independently and in class)? | .62 *** |
| 37. Encouraging self-directed learning? | .54 *** |
| 38. Respecting students' educated opinions? | .71 *** |

| | |
|--|---------|
| 39. Being available during office hours and/or by appointment? | .65 *** |
| 40. Providing helpful information during consultations? | .68 *** |
| 41. Keeping scheduled appointment times? | .51 *** |
| 42. Being accessible for questions prior to test dates? | .60 *** |
| 43. Providing additional information and resources for students? | .74 *** |
| 44. Having an approachable persona? | .69 *** |
| 45. Showing compassion and understanding to individual situations? | .73 *** |
| 46. Presenting key concepts clearly? | .70 *** |
| 47. Defining technical and ambiguous terms correctly? | .62 *** |
| 48. Providing new material and not simply reiterating textbook? | .40 ** |
| 49. Using examples to accentuate course material? | .55 *** |
| 50. Relating course material to 'real-life' application? | .40 ** |
| 51. Clearly explaining how each topic fits into the course? | .51 *** |
| 52. Developing abstract ideas and theories to connect material? | .68 *** |
| 53. Incorporating personal research into the lectures/seminars? | .17 |
| 54. Demonstrating interest and enthusiasm in the subject themes? | .54 *** |
| 55. Clearly communicating important 'take-home' principles? | .65 *** |
| 56. Fostering discussion within the area of study? | .64 *** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

Appendix B

List of the Characteristics Proposed by the Chair of the Psychology Department for Inclusion in the Target Standardized Questionnaire Construction

- Were course requirements clearly outlined?
- Was the workload appropriate for the course?
- Was the level of difficulty appropriate for the course?
- Was the grading fair?
- Was feedback timely?
- Was feedback helpful?
- Would you recommend this course to other students?
- Was the instructor well prepared?
- Was the instructor organized?
- Was the instructor audible?
- Was the instructor punctual?
- Did the instructor present key concepts clearly?
- Was the instructor available to students in office hours and/or by appointment?
- Was the instructor knowledgeable about the subject?
- Did the instructor encourage students to form their own judgements about the subject?
- Did the instructor respond adequately to questions raised by students?
- Would you take another course from this instructor?
- How would you rate your instructor's over all teaching?