**Why are you interested in the particular area of research?**

I’m interested in the field of noncognitive assessment because it’s an area that’s essential, rapidly growing in importance, yet has great room for improvement.

**The area of noncognitive assessments is important**

Noncognitive abilities are important in industry. Over the years, noncognitive variables have been shown to be able to predict a variety of work-related outcomes, including turnover (Salgado, 2000), task performance (Barrick & Mount, 1991; Hogan & Holland, 2003), and job satisfaction (Judge, Heller, & Mount, 2002). Due to their weak correlations with cognitive measures and significant incremental validity above and beyond intelligence measures, noncognitive tests have been considered an ideal supplement to cognitive ability tests, and widely used in the hiring processes.

Noncognitive abilities are also good predictors of educational outcomes. For example, among the five-factor mode (FFM) of personality, conscientiousness has been found to correlate strongly with academic performance (Poropat, 2009; McAbee & Oswald, 2013), and absenteeism (MacCann, Duckworth, & Roberts, 2009). However, unlike in industry, noncognitive ability are barely assessed for admission decisions. In a project that I just finished with Professor Brent Roberts last year, the PISA 2012 data was used to examine the prediction of math achievement and absenteeism from school by perseverance, controlling for SES and gender, across 9 major cultural groups consisting of all 68 participating countries and regions in PISA 2012. It was found that perseverance, especially the negative factor (including items such as giving up easily, and putting off difficult problems) of perseverance, was a great predictor of low math achievement and high truancy, and this relationship held across all 9 cultural groups.

Given the importance of noncognitive skills, the development of more efficient noncognitive assessments is one of the most fascinating research topics for me.

**There are still unanswered questions in this area**

To develop effective measures of noncognitive skills, there are a couple of important questions that need to be answered first, and these questions are the reasons for skepticism, criticism, or even attack from scientists and policy makers on the use of noncognitive assessments, especially in the field of education.

Modeling

When an IRT approach is adopted, what model to use for the analyses of self-reported noncognitive assessments is a major question. The debate has been going on for a while over which model suits personality data better, a dominance model that assumes the higher the trait level, the more likely the endorsement of the item, or an ideal point model that believes that highest endorsement derives only from the perfect fit between the trait and the statement. Fits of various dominance and ideal point IRT models have been examined and compared, and mixed results have been obtained. Based on my experience with my Master’s thesis, where both the SGR and the GGUM were applied to the Well-being and the Curiosity scales of the CPS (Wang, 2013), I think intermediate items are the key. An intermediate item is a type of item with a non-monotonic item characteristic curve (ICC), whose unfolding property can only be captured by an ideal point model. In my thesis, I attempted to compare the performance of the two models in DIF detection, but failed to do so due to the ill-conditioned matrices computed in GGUM2004 via the marginal maximum likelihood (MML) technique. However, what’s interesting was that under GGUM, the two scales actually exhibited less Differential Testing Functioning (DTF) than under SGR. This along with the fact that in our study the GGUM in general had better fit than the SGR, indicate that the GGUM may be useful in DIF detection, as long as we can figure out a better technique than MML to obtain accurate item parameter estimates.

Therefore, I would love to further investigate the suitability of different IRT models to response data of noncognitive assessments, administered not only in a single country, but across cultures. I’m also interested in reducing the discrepancy between researchers’ expectations for an item’s property and what the item’s ICC actually turns out to be (i.e., an item that was created to be intermediate may not actually turn out to have a non-monotonic ICC).

Fakeability

Unlike cognitive tests, the noncognitive ones are more often subject to faking, especially in a high-stakes setting where the test scores are used to make hiring or admission decisions. I think one key step to more fake-resistant noncognitive measures is to adopt a forced-choice format. This innovative response format that’s been used in the WorkFORCE™ Assessments will reduce biases induced by faking, and different response styles of participants of different cultural backgrounds, both of which are common criticisms of the Likert scale.

Therefore, I’m interested in the research on the development and scoring of noncognitive forced-choice tests (FCTs), as well as the possibility of assessing measurement equivalence with such a format, which is an area that’s basically blank.

To conclude, I’m passionate about the research area of noncognitive assessments and the WorkFORCE™ Assessments because it’s the future of more accurate prediction of a variety of work-related and education-related outcomes, and could be of great help for policy makers and employers, with many questions that remain unanswered.

**Why do you want to come to ETS?**

I want to come to ETS, because of its broad and deep expertise in item response theory and test development. It’s an organization full of the best talent in psychometrics, and offers the opportunity to collaborate with and learn from with many of the most prominent researchers in IRT. I believe that ETS is the best place for me to achieve my goals of adding to my knowledge and research experience in psychometrics. It would certainly be a great honor, and an eye-opening experience for me, to have an internship at ETS as a PhD student who’s passionate about exploring noncognitive assessments within the IRT framework.

Also, I appreciate the attitude that EST has towards research. ETS has never considered its long-term world-wide reputation in psychometrics as a reason to stop exploration. As far as I am concerned, ETS is an organization with the wisdom of a 70-year-old but the curiosity of a 7-year-old. As one of the world’s largest testing and assessment company, ETS is never satisfied with what it has achieved: from the paper-based GRE test to the computerized question-adaptive test until Aug 2011, to today’s multistage section-adaptive test, ETS is constantly exploring paths to a more accurate, secure, and user-friendly test. In the area of noncognitive assessments, ETS has impressed me with its rigorous efforts in applying the GGUM and the forced-choice format, which is still at its infancy, to the development and validation of the innovative WorkFORCE™ Assessments.

At ETS, research is the fuel of product development, and the products give feedback on the research quality, so the innovation and critical thinking will never cease. It’s the front line of both research and practice, where I will have the opportunity to apply what I have learned in lab to the real world, and to realize and in turn to reduce the gap between the two.

**What skills and knowledge will you bring to the internship?**

**Scientific software**

I have research experience with some of the most widely used IRT models (including 2PL, SGR, the GGUM) to study response data of personality tests. I have proficiency in using a variety of IRT-related software including BILOG, MULTILOG, GGUM2004, MODFIT, flexMIRT, and MCMC GGUM for item calibration, scoring, fit evaluation, and DIF analyses.

I started working with R in undergraduate school, and am proficient in R programing for simulation studies.

I’m also proficient in multi-level modeling in SAS (used in my 2016 PSPB paper with Professor Emily Grijalva), multi-group SEM in AMOS and Mplus (used in my PISA project with Professor Brent Roberts), and a variety of analyses in SPSS including regression analyses, random sampling, and data splitting and merging.

**Understanding of culture**

At the end of this semester, I’ll have finished my second seminar on cultural and social psychology with Professor Dov Cohen. From the seminars, I’ve gained a broad understanding of different cultures in the world and how culture influences almost every aspect of our lives. Some of the aspects that are closely relate to my research interest in cross-cultural noncognitive assessments are self-regards and self-views, social comparison, stereotypes and prejudice, and interpersonal relationships.

I believe my knowledge of culture will help me better interpret and explain results of noncognitive tests in a cross-cultural setting.

**Communication**

Verbal

I’m good at delivering oral presentations, and I don’t fear speaking in front of a crowd. When I give a presentation, my goals are to deliver the information clearly, and to make sure that everyone has a good time listening to me.

I did a good job giving my first-year presentation, where I, a first-year PhD student, presented the meta-analytic study on narcissism in front of more than 40 professors and more senior PhD students in the department. Instead of starting with a slide introducing research background, I captured the audience’s attention immediately by playing a short clip of Jenna Maroney from *30 Rock*, which vividly described to the audience with various research backgrounds what a grandiose narcissist looks like.

I gained my confidence and competence in public speaking also from my experience of teaching undergraduate students for the past 7 semesters. I’ve taught both labs of 20 students and lectures of 60, and have enjoyed all of them.

Written

I’m skilled in conveying information in a professional and timely manner. On average, I deal with more than 20 emails a day, including emails from the department, advisors, projects collaborators, and undergraduate students. The content of the email covers school affairs, academic research, course requirements, grades and so on. I’m usually able to respond to emails within 2 hours. More than three years of dealing with various types of professional email has prepared me well with daily professional written communication.

Interpersonal

The experience of running discussion sessions, participating seminars, as well as interaction with my undergraduate students has strengthened various interpersonal skills of mine. For example, I’ve held brainstorm and discussion sessions in a lab of 8 people for item writing for a personality assessment development project, where I was responsible for motivating discussions, and steering the conversations back to the topic from time to time. I needed to listen carefully to lab members’ opinions, ask relevant and engaging questions, and respect the members by not interrupting them, even when the conversation became heated. However, when I’m interacting with my students, I focus on understanding students’ problems, and in turn crafting informative and thoughtful responses.

**What contributions do you believe you can make to a project in that research area?**

**Knowledge and skills necessary for a project**

I have almost 4 years of research experience relevant to noncognitive assessments, which is my primary research interest.

Master’s thesis

My Master’s thesis adopted an IRT approach to studying DIF of the Well-being scale and the Curiosity scale from the Comprehensive Personality Scale (CPS; Wang, 2013), using data collected from China and the United States. A dominance IRT model (i.e., SGR) and an ideal point model (i.e., the GGUM) were applied to the data, and within the Null Hypothesis Significance Testing (NHST) paradigm, the constrained and the free baseline approaches along with the log-likelihood ratio (LR) test were adopted to examine DIF. Also, a DIF effect size measure (Nye, 2011) was used in order to obtain DIF magnitude, and compensate for the oversensitivity to large samples of the NHST paradigm.

Thanks to this project, I am now proficient in using MULTILOG, GGUM2004, and MCMC GGUM for item calibration, and in R programing for a variety of IRT analyses, including running packages for DIF detection (e.g., “lordif”, “difR”), data generation based on different IRT models, writing functions that automatically generate syntax for MULTILOG and GGUM2004, and calling from R external IRT programs to run automated analyses such as the constrained baseline and the free baseline modeling. I’m also very familiar with MODFIT, the Excel macro for computing the fit for IRT models, and I am able to quickly get model fit computed from it and give accurate interpretation of the results for various IRT models (e.g., 2PL, 3PL, SGR, GGUM…). I have no problem evaluating item features and qualities based on the ICCs, as in the course of working on my thesis, I saw and interpreted over a hundred of them, for the purpose of recognizing non-discriminating items, unfolding items, and negative items. Last but not least, I am also capable of transforming MATLAB code to R code, as I did with the MATLAB syntax for the DIF effect size measure.

The PISA project

In a project that I just finished with Professor Brent Roberts last year, I had the opportunity of using CFA, a different approach than IRT to study measurement equivalence across cultures. The project used the PISA 2012 data to examine the prediction of math achievement and absenteeism from school by perseverance, controlling for SES and gender, across 9 major cultural groups consisting of all 68 participating countries and regions.

This project gave me the opportunity to get proficient in multigroup structural equation modeling (SEM), in both Mplus and AMOS, as well as carrying out data splitting and merging, random sampling, and factor analyses in SPSS.

I’m not only good with numbers and programs, I’m also good with searching, organizing, and processing information in other areas. For example, on the PISA project, I was responsible for grouping the 68 countries into 9 major cultural groups, partially based on Saucier and colleagues’ (2005) grouping of 38 countries. However, I felt that their way of grouping, which considered only geographic proximity, was not good enough for forming groups that are representative in terms of culture, so I took a step further. I referred to other sources including The United Nations Statistics Division (UNSD) website, and Wikipedia, and eventually came up with a grouping strategy that took into consideration not only proximity, but also other factors such as history, religion, politics, and economy conditions.

The Conscientiousness scale development project

In another ongoing project with Professor Brent Roberts, we are trying to develop a new Conscientiousness scale based on the Chernyshenko Conscientiousness Scale, but with more facets included, and fewer items in each facet.

In order to create a large item pool, I held discussions in the lab, where all members wrote some new items for each facet, talked about them, and decided if they were clear and accurate enough to be kept in the item pool. Sometimes, I needed to be the one to initiate the discussion, and motivate everyone to participate, especially at early morning meetings, while sometimes I needed to steer the discussion back to the topic. While the discussion went on, I was also the responsible for putting down the new items along with lab members’ comments on them. This experience equipped me with the communication skills necessary for a group project, and the ability to select and organize information. Therefore, when there’s a new project, I will have no problem leading discussion or brainstorm sessions, while putting down important information in an orderly fashion for later.

Qualities

Apart from the skills I’ve acquired through my research experience, I believe that some of the qualities of mine will contribute to the project as well.

Ever since I came to UIUC, I’ve worked with a variety of more senior PhD students and faculties, both locally and remotely. I’ve had a great time collaborating with all of them, and have learnt a great deal from them, but being a team player does not mean that I’ll just conform blindly.

For example, during my first year, when I was coding in MULTILOG for a DIF assignment, I was not sure if I should put down 1 or 2 for the number of groups, because I had seen both. A senior student in the lab told me to put down 1 as that was how he and other people he knew had been doing it. He put it as it was common sense. However, I was not persuaded at all by the reason “everyone else is using 1”. After consulting other more experienced researchers and looking it up in more papers and research reports, I finally confirmed that the correct group number should be 2, although the code for equating item parameters implicitly identified two groups for MULTILOG.

I informed that senior student about the correct group number, and fortunately, he hadn’t published any of those projects yet.

Without sufficient evidence, I won’t take any conclusions for granted, regardless of where it’s from, or whether it’s considered obvious by many. If I find something confusing, I’ll ask.

I’m also a persistent explorer.

Until early last year, there was a problem that the “shell” function in R to call GGUM2004 through a .au3 file was only working on two of all the computers in our lab, including our personal computers. This is a function that’s essential to automated GGUM analyses such as simulations, and baseline modeling for DIF. To avoid waiting in line for computers to run analyses, I decided to solve this problem once and for all. I was confident that I could get it done, because with a close check, I believed the problem was the syntax of the .au3 file, so I spent five days learning about the basic .au3 syntax, and finally located the source of the problem – the file path, and the missing cmd file. After fixing that, I also added two lines of code, so that now the cmd window of GGUM2004 closes automatically to let the simulation continue if there is a singular matrix error, whereas under the old code, the window would stay open and the automated process would just stop until someone closed the window manually.

I shared my findings and the code with the lab, and now everyone in the lab can easily get the “shell” command to run on any PC with GGUM2004 installed. I am perseverant in solving problems encountered in the course of research, and I believe this quality of mine will contribute to other projects, especially those about exploring alternative scoring for the forced-choice response format, which is a relatively new area with a lot of unanswered questions.

**References**

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis.*Personnel Psychology, 44*(1), 1-26. Retrieved from <http://search.proquest.com.proxy2.library.illinois.edu/docview/617946789?accountid=14553>

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective.*Journal of Applied Psychology, 88*(1), 100-112. doi:http://dx.doi.org.proxy2.library.illinois.edu/10.1037/0021-9010.88.1.100

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis.*Journal of Applied Psychology, 87*(3), 530-541. Retrieved from <http://search.proquest.com.proxy2.library.illinois.edu/docview/619737885?accountid=14553>

MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of conscientiousness.*Learning and Individual Differences, 19*(4), 451-458. Retrieved from <https://search-proquest-com.proxy2.library.illinois.edu/docview/621948288?accountid=14553>

McAbee, S. T., & Oswald, F. L. (2013). The criterion-related validity of personality measures for predicting GPA: A meta-analytic validity competition.*Psychological Assessment, 25*(2), 532-544. Retrieved from <https://search-proquest-com.proxy2.library.illinois.edu/docview/1285630367?accountid=14553>

Nye, C. D. (2011). *The development and validation of effect size measures for IRT and CFA studies of measurement equivalence.*Available from PsycINFO. (1269433733; 2012-99220-321). Retrieved from <http://search.proquest.com.proxy2.library.illinois.edu/docview/1269433733?accountid=14553>

Poropat, A. E. (2014). A meta‐analysis of adult‐rated child personality and academic performance in primary education. *British Journal of Educational Psychology*, 84(2), 239-252. Retrieved from <https://search-proquest-com.proxy2.library.illinois.edu/docview/1548790325?accountid=14553>

Salgado, J. (2002). The big five personality dimensions and counterproductive behaviors.*International Journal of Selection and Assessment, 10*(1-2), 117-125. Retrieved from <http://search.proquest.com.proxy2.library.illinois.edu/docview/619877040?accountid=14553>

Wang, W. (2013). *A Bayesian Markov chain Monte Carlo approach to the generalized graded unfolding model estimation: The future of non-cognitive measurement.* Available from PsycINFO. (1676371094; 2015-99080-541). Retrieved from <http://search.proquest.com.proxy2.library.illinois.edu/docview/1676371094?accountid=14553>