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

I’m interested in the filed of noncognitive assessments because it’s an area that’s essential, promising, yet still 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), organizational citizenship behavior (OCB; Borman, Penner, Allen, & Motowidlo, 2001), counterproductive work behavior (CWB; Donnellan, Spilman, Garcia, & Conger, 2014), leadership (Judge, Bono, Ilies, & Gerhardt, 2002), 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 intelligence 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 (Costa & McCrae, 1992; Goldberg et al., 2006), conscientiousness, either the broad dimension or the narrow facets of it, has been found correlate strongly with AP (e.g., Poropat, 2009; McAbee & Oswald, 2013; Morris & Fritz, 2015), 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 a core issue that fascinates me.

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

However, to develop better measures for noncognitive assessments, 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 filed 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 whether a dominance model that assumes the higher the ability, 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 ability and the statement, is more appropriate for such assessments. Fit 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 Comprehensive Personality Scale (Wei, 2013), and I think intermediate items is the key, a type of item that can only be modelled well by the ideal point models, but not the dominance models. In my thesis, I also attempted to compare the performance in detecting DIF of the two models, but failed to do so with the GGUM, due to ill-conditioned matrices given 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 results that the GGUM in general had better fit than the SGR in my thesis, are indicating 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 estimates.

Therefore, I’m very interested in further investigating the suitability of applying different IRT models to response data of noncognitive assessments, administered not only in a single country, but across cultures. I’m also very interested in the 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 having a non-monotonic ICC).

Fakeability

Unlike cognitive tests, noncognitive ones are easy to fake, especially in a high-stake setting where the test scores are used to make hiring or admission decisions. And I think one key step to increase the tests’ resistance to faking is by adopting a new response format that’s hard for respondents to fake. This innovative response format adopted in the WorkFORCE™ Assessments will see fewer biases induced by faking in a high-stake situation, and by different response styles of participants due to their cultural backgrounds, both of which are very commonly seen criticism of Likert scales. Therefore, I’m very interested in the research on the scoring of forced-choice noncognitive tests (FCTs), which is the essential prerequisite for a wide acceptance and application of such a response format, and I would also like to explore the possibility of assessing measurement equivalence with such a format, which is an area that’s basically blank.

To conclude, I’m interested in the research area of noncognitive assessments and the WorkFORCE™ Assessments because I believe that they are the future of more accurate prediction of a variety of work-related and education-related outcomes, and can be of great help to employers and policy makers.

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

I want to come to ETS, because it’s among the first to invent item response theory, and has no doubt long been one of the ultimate experts in the area. Plus, it’s an organization full of the best talent in the area, and has a lot of collaboration with the most prominent researchers in IRT. It would be a great honor, and an eye-opening experience for me, if I could come to work at ETS as an intern as a PhD student, who’s passionate about exploring noncognitive assessments within

the IRT framework, I believe that ETS is the best place for me to achieve my goals of adding to my knowledge and research experience in psychometrics.

Besides growth in my knowledge and skills, I also appreciate a lot the attitudes of EST towards research. ETS has never considered its long-term world-wide reputation in psychometrics as a reason to stop the exploration. It’s 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 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. This spirit of untiring investigation is what I value a lot about ETS. At ETS, research is the fuel of its products, and the products in turn examine the quality of the research, and innovation and critical thinking will never cease. It’s the frontline of both research and practice, and thus it will the best place for me to apply what I have learnt in the lab and in classroom to the real world.

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

**Item response theory and programming**

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, test scoring, model-data fit evaluation, and DIF analyses.

I started working with R in undergraduate school, and have proficiency of using R for simulation programming, and automated programming for the constrained and free baseline approaches to DIF analyses.

I’m also proficient with 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).

**Understanding of different cultures**

I’ve taken two of Professor Dov Cohen’s seminars on cultural and social psychology, and have a broad understanding of how culture influences almost every aspect of our lives, some of the aspects that relate to my research interest in cross-cultural personality assessments are self-regards (i.e., how people view themselves), comparison, socially-desired values, and so on.

I believe my knowledge of culture will likely to help me with understanding and explaining results obtained from cross-cultural personality assessments, and it will also give me more confidence working in the diverse environment at ETS.

**Communication skills**

Verbal

I’m proficient in delivering oral presentations, and I don’t fear speaking in front of a crowd. When I give a presentation, my goals are delivering clearly the information I want to deliver and trying my best to make sure that everyone has a good time. I did a good job on my first-year presentation, where I, as a first-year PhD student, presented the meta-analysis on narcissism in front of more than 40 professors and more senior PhD students in the department. Instead of starting directly with a slide introducing research background, I attracted the audience attention by showing a short clip of Jenna Maroney from 30 Rock, which vividly described to the audience what a grandiose narcissist is like.

I gained my confidence and competence in public speaking also from my experience of teaching undergraduate students for 8 semesters so far. I’ve taught both labs of 18 students and lectures of 60. When I’m giving a talk, I hold an honest attitude, so that the possibility of being judged or of not knowing the answer to a question from the audience never bothers me.

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 usually am able to respond to emails within 2 hours. Four years of dealing with various types of professional email has prepared me well with everyday professional written communication.

Interpersonal

The experience of running discussion sessions, participating seminars, as well as four years of 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 lab members’ opinions, ask relevant and engaging questions, and respect the members by not interrupting them, even when the conversation gets heated. When I’m talking with my students, however, my major tasks will be first listening to them closely, and then 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 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), 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 was used in order to obtain the magnitude of DIF, and compensate for the oversensitivity of NHST paradigm to large samples.

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 start 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 them for various IRT models (e.g., 2PL, 3PL, SGR, GGUM…). I have no problem reading the ICCs, the visual aids for item evaluation, as in the course of working on my thesis, I saw and interpreted over a hundred of them, for the purpose of recognizing non-discriminant 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 DIF effect size measure syntax.

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 in PISA 2012.

Thanks to this project, I’m now proficient with multigroup structural equation modeling (SEM), in both Mplus and AMOS, as well as carrying out in SPSS data splitting and merging, randomly sampling, and factor analyses.

In PISA 2012, there is a wide variety of sample sizes, ranging from 293 for Liechtenstein to 33806 for Mexico, so in order to avoid potential biases brought by the drastically different sample sizes, we decided to randomly sample 500 people from each country/region for the analyses (i.e., Liechtenstein was dropped from the analysis). This was done in SPSS, where I performed a variety of basic analyses to prepare the data for SEM, including data split and merging, factor analyses, correlations by and across groups. ~~Through this project, I gained hands-on experience with modeling measurement equivalence using large-scale data with AMOS, Mplus, and SPSS.~~ I also was responsible for grouping the 68 countries into 9 major cultural groups, which was done partially based on Saucier and colleagues’ (2005) grouping of 38 countries, but I felt that their way of grouping, which considered only geographic proximity, was not enough for the form of groups that are representative and discriminate from each other in terms of culture, I took a step further, and consulted The United Nations Statistics Division (UNSD) website, as well as Wikipedia, and eventually came up with a grouping strategy that took into consideration not only proximity of the countries, but also other factors such as cultural background, political systems, and economy conditions. I’m not only good with numbers and programs, I’m also good with searching, organizing, and processing information that’s closely related to the society and the world.

In another ongoing project with Professor Brent Roberts, where the goal is to come up with a new version of the Chernyshenko Conscientiousness Scale, with more facets covered, but fewer items contained in each facet. In order to develop a new scale, we needed to first create a relatively large item pool, so I would hold discussion session in the lab, where every member came up with new items for each of the facets, and talk about them and decide 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 during meetings early in the morning, while sometimes I needed to steer 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 of information selection and organization. Therefore, when there’s a new project, I will have no problem leading discussion or brainstorm sessions within the group, while putting down important information in an orderly fashion for later.

Qualities and spirits

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

One thing I’ve learnt from taking and teaching my social psychology classes is the danger of conformity. I’m a team player, but I don’t conform.

Ever since I got into graduate school, 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 that doesn’t mean I’ll hesitate asking questions when something is not right. For example, during my first year, when I was coding in MULTILOG for a DIF study, 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’s what he and some others he knew had been using. He put it as it was common sense. I was not persuaded at all by the reason that “everyone else is using 1”. After consulting with other more experienced researchers and looking it up in papers and research reports, I finally confirmed that the correct group number was supposed to be 2, instead of 1, although the code for equating item parameters implicitly identified two groups for MULTILOG. I privately informed the student in the lab about the correct group number, and fortunately, he hadn’t published any of those projects yet. What was interesting was that both sides I consulted thought their answer was the obviously correct one, and had never given it too much thoughts before I asked. I think what I can contribute to a project is that I don’t take any conclusions for granted, regardless of whether it’s from a colleague or a supervisor, and whether it’s a considered by many an obvious one. If I find something confusing, I’ll ask.

I’m also an explorer, and I like to share my findings. In the same DIF project mentioned above, when I was running the baseline models in R under GGUM, I needed to call an external GGUM2004 program from R. To start the program, I needed to use the “shell” command in R to first call a .au3 file written by a computer sciences student several years ago, which initiates the command file of GGUM2004. This .au3 file is handed down in the lab by former students, but the problem with it we’ve been having recently in the lab is that this “shell” command is working only on the computer of one student in our lab, and he had no idea why, and the technical group on campus was of much help, either. This student kindly offered his computer to me when I needed to run a simulation study with the “shell” command. As grateful as I was, I said no. This was because I wanted to solve this program problem once and for all, rather than avoiding it, and I was confident that I could solve this problem, because with a close check, I believe the problem lies in the syntax of the .au3 file used to call GGUM2004. Next, I spent 3 days learning about basic .au3 syntax, and finally located the source of the problem – the file path is wrong, and the cmd program is missing from the assigned folder. I also added two lines of code that enables the cmd window of GGUM2004 to close to let the simulation continue if there’s a singular matrix. Under the same circumstance, with the old .au3 code, the window won’t close itself and the automated process will be paused. I shared my findings and the code with the lab, and now everyone in our 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 personality projects, especially those about exploring alternative scoring for the forced-choice responses, which is a relatively new area with a lot of unanswered questions and mysteries.