PILOT EVALUATION OF THE COMPUTER-BASED ASSESSMENT OF NON-COGNITIVE ATTRIBUTES OF HEALTH PROFESSIONALS (CANA-HP)

by

SARA MAHER DISSERTATION

Submitted to the Graduate School of

Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

2020

MAJOR: EDUCATION AND EVALUATION

Approved By:

Shlomo Sawilowsky 9/22/2020
Shlomo Sawilowsky 9/22/2020 (Sep 22, 2020 12:38 EDT)

Advisor Date

Barry S Markman
Barry S Markman (Sep 22, 2020 13:56 EDT)

Peter D. Frade
Peter D. Frade (Sep 22, 2020 14:16 EDT)

DEDICATION

I dedicate this dissertation to three sets of unique individuals who helped to shape my life and my quest to never stop learning.

First, to my mom and dad.

You gave me the gift of not being afraid to try.

You fostered my love of school which has never waned.

Second to my grandmother Merle.

You taught me to be kind, to laugh, and always make time for fun!

Finally, to my husband Jim.

You have given my wings to fly.

There is nothing I cannot try without you by my side.

ACKNOWLEDGMENTS

I wish to express my deepest thanks and sincere gratitude to my dissertation committee members. This could not have happened without you. Dr. Shlomo Sawilowsky, you provided unwavering support, sage guidance, and a love of psychometrics which inspires me. You hold the world's record for fastest editing time! You are the reason this work moved forward to quickly. Dr. Barry Markman, you allowed me to venture into the world of consulting through my directed study. Your guidance working with others provided invaluable insight for my future work mentoring faculty and students. Dr. Whitney Moore, your class has become a foundation for most of the analyses I use in my own research projects. I still refer to your class notes for every research study I conduct. Dr. Peter Frade, you have been my mentor and role model since I came to Wayne State University. You are the greatest advocate for faculty I know, and I am blessed to work with you.

I also wish to acknowledge my mentors from Oakland University. Dr. Stiller, you gave me the inspiration to submit my first manuscript, thus beginning my journey into academia. Dr. Thompson, you were the best role model and mentor a new faculty member could ask for. None of the advice you gave me, ever proved wrong, even when it came to being an administrator.

Finally, I need to thank my life long editor and partner in this journey, my husband Dr. James Maher. You read papers, listened to speeches, and choose to forego family plans so I could study. You have made it through my last three degrees, and even though I keep saying this will be my last. This time I mean it!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 – HOLISTIC ADMISSIONS

Health Care Discrepancies

In an era of health care reform, one area under scrutiny has been diversity in the health care workforce. Millions of newly insured patients, many from underserved areas, have begun seeking health care services not previously available to them. By the year 2030, it has been projected children of racial/ethnic minorities would account for over 50% of the population under the age of 18 in the United States (Meadows, 2014). To meet the needs of complex and underserved patient populations, Artinian et al. (2017) identified health care professionals must possess diverse backgrounds, qualities, and skill sets.

A Healthy People agenda was developed by a task force within the U.S. Department of Health and Human Services (USDHHS) in 2000, and focused on reducing health inequity (Meadows, 2014) or health equity. “Achieving health equity required valuing every person equally, with focused and ongoing societal efforts to address avoidable inequalities, historical and contemporary injustices, and the elimination of health and health care disparities” (Meadows, 2014, p. 2). To achieve health equity, health care professionals must be educated in environments which value diversity, and those selected for admission into such programs should possess background, skills, and other qualities to enable treatment of patients from diverse backgrounds with complex needs (Artinian, et al., 2017; Meadows, 2014).

Strategies were sought to diversify the student population, with an overall aim to ultimately diversify health care workers (de Visser et al., 2018; DiBaise et al., 2015; Kalsbeek, 2013). To achieve a diverse workforce can be challenging. Shields (2010) claimed “to diversity our workforce, it will mean broadening our sense of fit, and

acknowledging a wider range of knowledge, skills, and attributes" (p. 59). In theory, a diverse workforce would allow patients to be treated by practitioners of similar backgrounds which could increase comfort level, model healthy behavior, and help to avoid inequalities in the provider - patient relationship. A more diverse workforce was shown to improve patient satisfaction, improve health access and equity, and increase the recruitment of minorities into the health professions (DiBaise et al., 2015). Patients have a tendency to select health care workers who have similar ethnic backgrounds to the patient (Gould, 2014).

Despite the recognized benefits of a diverse health care workforce, the proportion of under-represented minorities applying for admission into health professions programs remains low. For example, it was shown in the 2017-2018 Aggregate Program Data Fact Sheet (Chana, 2017-2018) disparities existed in the percentage of students accepted into physical therapy school. Among the accepted physical therapy students, the following ethnic / racial distributions were observed; 3.26% African American, 0.43% American Indian / Alaskan Native, 8.21% Asian, 0% Asian / Pacific Islander, 75.9% Caucasian, 6.29% Hispanic / Latino, 2.42% of 2+ origins, and 3.07% who declined to identify a race (Chana, 2017-2018). Similar results were reported in the 2017-2018 Occupational Therapy Annual Data Report with even lower distributions rates as follows: < 1% American Indian, 7% Asian, 3-5% African American,< 1% Pacific Islander, 80-85% Caucasian, and 5-8% unspecified (Harvison, 2017-2018). Among physician assistant programs, graduates had the following racial / ethnic distribution in 2013; 2.8% African American, 0.2% American Indian, and 7.4% Hispanic / Latino (DiBaise et al., 2015). Similar trends were seen in nursing, with only 27% of students coming from minority

backgrounds (Gould, 2014). Without an adequate pool of graduates, it is impossible to increase the number of multi-cultural graduates in the workforce.

For students who choose to apply, the admission processes for many health professions programs were fraught with inequality, with criteria for successful admission related to academic skills such as high overall college GPA, high science GPA, and high GRE scores. For example, the average cumulative grade point average of students who applied to physical therapy school was 3.59, and most admission committees required a minimum of 3.0 cumulative GPA to even score an applicant (Chana, 2017-2018).

High GPA's were a standard for all health professions, and pre-requisite class requirements often included required concentrations in math and science. With large numbers of students competing for few seats, preference was often given to the advantaged, those students who were able to successfully navigate the standardized test. As Howe (1997) noted, "educational testing fails to take into account educational inequalities experienced by children both in and out of schools" (p. 101). These educational inequalities were so distinct "even individuals who are talented but disadvantaged by social situations do not perform well on examinations" (p. 91), and "Certain groups are disadvantaged by educational testing, and they may receive different opportunities as a result of the testing" (p. 92).

In admissions to health care professions, this may mean qualified individuals from marginalized groups in society may not be afforded the opportunity to become a health care professional due to having low GPA or GRE scores. The challenge for health care educational programs has been to identify methods to admit students who better represent all patients requiring medical treatment, patients who come from a broad

spectrum of backgrounds including individuals of differing races, social classes, sexual orientations, languages, religions, and countries of origin (Shields, 2013).

Among physician assistant programs, the greatest barriers to admission into a program were identified as the following; legal issues (state policies, court decisions, state legislation on affirmative action), educational preparation (pre-requisite classes, high school attended), sociocultural factors (lack of role models, peer/community support), financial/economic issues and recruitment and admission factors (DiBaise et al., 2015). Often such barriers inhibited minority applicants from getting accepted into health care professions.

Holistic Admissions

A strategy employed to diversify student admissions was the use of holistic review, a “flexible, individualized method of assessing an applicant’s attributes, experiences, and academic metrics to determine how the individual might contribute as a student and future health care professional” (Artinian, et al., 2017, p. 65; Witzburg & Sondheimer, 2013, p. 1565). A holistic admission may be considered a broad-based admission which looks beyond the academic preparation each applicant brings to the admission process (Kalsbeek, 2013). Holistic assessment focuses on the non-cognitive attributes of a candidate, rather than the traditional cognitive attributes which have been theorized to be measured by tests such as the Graduate Record Examination (GRE) and Medical College Admission Test (MCAT).

Non-cognitive attributes have varied based upon a school’s mission, local context within a program, as well as the state in which a program was located (Artinian, et al., 2017). Some of the non-cognitive variables which have been used in admission criteria

included: commitment to service activities, cultural sensitivity, empathy, capacity for growth, emotional resilience, strength of character, interpersonal skills, and curiosity / engagement (Witzburg & Sondheimer, 2013). In addition, some admission committees have used a screening process to identify non-academic criteria for admissions which included: first generation status, socioeconomic status, race or ethnicity, foreign language ability, gender, experience with disadvantaged populations, origin in a community with health profession shortages, origin in a community targeted by the school, and any other attribute specific to a school / program mission, geographic context, or workforce need (Artinian, et al., 2017).

Purpose of the Study

Although a number of tools have been developed for cognitive performance, a common tool has yet to be identified which can effectively screen the non-cognitive attributes of applicants seeking admission into different heath care programs. The purpose of this study, therefore, was to develop an admission tool which could effectively screen non-cognitive attributes of applicants seeking admission into one of four health care professions: (1) nurse anesthesia, (2) occupational therapy, (3) physician assistant, and (4) physical therapy.

The *Computer-based Assessment of Non-Cognitive Attributes of Health Professionals* (CANA-HP) is a methodology which is being developed to efficiently screen non-cognitive attributes of a variety of health care professions. Although it has been refined based upon several tools which have been studied beginning in 2004, the assessment introduces several new attributes which have not been previously tested. First, the tool will compare open-ended questions with rank-order questions or best choice

questions. Second, all questions will be delivered in a computer-based format with time limits. Finally, the questions have been developed specifically for use with a wide variety of health care profession applicants. Therefore, initial analysis of these questions was not completed prior to this study.

Study Questions

The broad research question for this study was "What are the psychometric properties of a *Computer-based Assessment of Non-cognitive Attributes of Health Professionals* (CANA-HP)?" Three specific questions were delineated, focusing on different aspects of the CANA-HP:

1. What is the CANA-HP instrument reliability (internal consistency & interrater) within each station (rater) and inter-station (station)?
2. Does the CANA-HP measure attributes of non-cognitive variables as demonstrated by low construct validity scores when correlating the CANA-HP to traditional assessments reported to measure cognition (e.g. pre-admission GRE and GPA)?
3. Does analysis reveal differences between groups based upon gender, ethnicity, Pell-grant status, family history of college, or socio-economic differences?

Assumptions

The non-cognitive attributes included as part of the CANA-HP may or may not be applicable across different health professions. This study assumes there are certain non-cognitive attributes which are universally desired by any health care professional. To help clearly define these attributes, content experts from each health profession were consulted at the same time the CANA-HP was developed.

It is also assumed construct validity of non-cognitive variables can be determined through low correlations with items thought to measure cognition. The statement implies cognitive thinking is minimally related to non-cognitive variables and the tools selected to measure cognition (GRE, GPA) clearly measure the latter construct. Finally, it is a premise to assume admission committees for health professional programs are interested in examining non-cognitive variables in the admission process.

Limitations

The study is limited to applicants into one of four health care programs at Wayne State University. The aim is to show psychometric properties of the CANA-HP across different disciplines. Because there are a number of health professions at Wayne State University who annually admit students in different health programs, this question can be addressed by applicants at this university.

The sample is limited to only those applicants who come to the Wayne State University for an interview. Because verification of identify is a concern, only individuals who can be verified through picture identification will be allowed to participate in the study. This limits the ability to generalize the use of the CANA-HP with all applicants who seek admission into one of the four Wayne State identified health profession programs (nurse anesthesia, occupational therapy, physician assistant, and physical therapy).

Definition of Key Terms

Computer-based Assessment of Non-cognitive Attributes of Health Professionals (CANA-HP): The CANA-HP is a measurement methodology developed to contain 12 situational judgment tests (SJT). Six of the SJTs were open-ended scenarios and six were formatted in a traditional ranking or best answer format. The SJTs were presented in a

computerized format, with an a 60 minute time frame allowed for the applicant to read each situation presented and answer any subsequent question(s).

Holistic Admissions: A flexible, individualized method of assessing an applicant's attributes, experiences, and academic metrics to determine how the individual might contribute as a student and future health care professional (Artinian, et al., 2017; Witzburg & Sondheimer, 2013).

Non-cognitive attributes / variables: Attributes of an individual which are not related to traditional verbal and quantitative areas typically measured by standardized tests. These attributes include, but are not limited to areas such as personal and social dimensions, motivation, adjustment, ethics, critical thinking, and knowledge of health care situations (Sedlacek, 2017)

Situational judgment test (SJT): A hypothetical written scenario or situation is presented and the reader is asked any number of questions to probe how the individual responds to the situation presented (Patterson, Zibarras, & Ashworth, 2016; Shipper, et al., 2017) Open-ended SJTs contain the scenario with one or more broad questions which a reader answers in essay form. Traditional SJTs contain a situation or scenario followed by question(s) written in multiple choice, ranking or best single answer format (Table 1).

Table 1.

Examples of a Situational Judgment Test Showing Different Response Formats

Multiple choice	**Ranking**	**Best single answer**
You review the chart of a patient and determine the patient may be taking medication which may have potential dangerous interactions which could	You are treating a patient who has previously been diagnosed with cancer. Prior to starting your treatment, the patient leans toward	A patient has been prescribed painkillers to help during the first three days following surgery. The patient expresses pain killers are not

harm the patient. The staff nurse challenges your decision to call the attending physician.

Choose the THREE most appropriate responses in this situation.

a. Instruct the nurse to immediately call the physician.
b. Discuss with the nurse the reasons for her disagreement.
c. Ask a senior colleague for advice.
d. Complete a clinical incident form.
e. Arrange to speak to the nurse later to discuss your working relationship
f. Write in the medical notes your thoughts about the medication error and indicate the staff nurse declined to call the physician,
g. Review the case again.

you and quietly asks “Is my cancer back”?

Rank in order of appropriateness the following actions in response to this situation.

a. Explain to the patient the cancer has returned.
b. Reassure the patient he/she will be fine.
c. Explain to the patient the results are not back, but you will speak to him/her when the results are in.
d. Inform the patient you will look up the results of the test and have a colleague discuss.
e. Invite the patient to join you and a senior colleague in a quiet room to discuss a cancer diagnosis and explore fears.

good for overall health, and the patient is opposed to taking them.

What is the BEST way for you to react to the patient refusal to take the prescribed medication?

a. Ask the patient if he/she knows something else to relieve pain.
b. Give the scientific evidence as to why painkillers will work.
c. Agree with the patient to avoid pain killers for now and try other treatment methods.
d. Tell the patient he/she needs an attitude change to be more open to all treatment options.

CHAPTER 2 – LITERATURE REVIEW

Cognitive Versus Non-cognitive Attributes

According to Sternberg (1985), intelligence consists of three subsets, or a triarch, rather than a single ability. The contextual or practical subset was identified as the ability to 1) adapt to a current environment, or 2) select a better environment than the one an individual is operating in, or 3) shape the current environment to make it a better fit for the individual. The experiential or creative subset was demonstrated when an individual interprets a novel task or situation or is in the process of automatically responding to a task or given situation. The componential or analytical subset involved the ability of an individual to interpret information hierarchically in well-defined and unchanging contexts (Sternberg, 1985).

According to Kalsbeek (2013), standardized tests measure only one subset of intelligence, the componential / analytical subset. Analytical or cognitive atrributes were traditionally screened during the admission process into a health care field through a number of standardized assessments such as the GRE and GPA (Kalsbeek, 2013). However, experiential / creative and contextual / practical intelligences were thought to be the methods individuals from non-traditional backgrounds used first to learn. Through the use of these latter subsets of intelligence, individuals from non-traditional backgrounds began to move componential / analytical intelligence to the forefront of their learning (Kalsbeek, 2013). The question becomes how to test creative and practical intelligence because they are not traditionally measured through standardized testing such as the GRE, SAT, etc. One hypothesis is the two subsets may be best assessed by other variables such as non-cognitive attributes, often used as part of the holistic

admission process.

Non-cognitive attributes, as previously defined, focused on characteristics of individuals beyond traditional educational testing. Non-cognitive variables were useful for all students as “they provide viable alternatives in assessing the abilities of people of color, women, international students, older students, students with disabilities, LGBTQ students, or others with experiences which are different from those of young, White, heterosexual, able-bodies, Eurocentric males in the United States” (Sedlacek, 2017, p. 28).

A concern with using non-cognitive variables as part of the admission process has been the impact on student outcomes. For example, de Visser et al. (2018) compared two independent cohorts of students, one selected with traditional cognitive variables and the other selected with non-cognitive variables. The dropout rate was highest in the non-cognitive group. The non-cognitive admission cohort, however, had a higher percentage of students who received the maximum grade for first year nursing school and had higher grade point averages for practical clinical courses in the 3rd year of the program. There were no statistically significant differences in GPA during the 1st and 2nd years in the program (de Visser et al., 2018).

Stratton and Elam (2014) examined the predictors of underperformance during the first year of medical school. Results indicated underperformers included students over 31 years of age, African American students (the largest proportion of underperformers), students who had significantly lower GPAs at the undergraduate level, students who entered medical school via an accelerated track, or applicants who were admitted with a non-unanimous decision by the admission committee. Academic underperformers were

found to be significantly less conscientious (Stratton & Elam, 2014). In general, neither cognitive nor non-cognitive variables predicted an applicant's success in a heath profession.

There are limitations to using either a cognitive or non-cognitive approach for candidate selection into a health profession program. Students underperformed in medical school, for example, both in cognitive and non-cognitive reasons, making it difficult to determine which causal factor contributed most to a student who was not successful. Another limitation is non-cognitive variables are hard to test through pre-screening and definitive constructs have not been established. For example, clinical reasoning or the process by which a health care professional assesses a patient, has not been previously assessed with a standardized tool. However, non-cognitive variables have been studied by a number of authors beginning as early as 2000 with the multiple mini-interview.

Non-cognitive Variables Assessed with the Multiple Mini-Interview (MMI)

The Multiple Mini-Interview (MMI) was an assessment process developed for medical school admission by Eva et al. (2004). The MMI was designed as a structured selection method where applicants rotated through a series of ten stations designed to test non-cognitive attributes. Candidates were not expected to have specialized knowledge, rather candidates were expected to think logically through a topic and communicate with an interviewer effectively.

Each station involved a one-on-one discussion between the interviewer and candidate with structured questions in four domains: (1) critical thinking, (2) ethical decision making, (3) communication skills, and (4) knowledge of the health care system

(Eva et al., 2004). Three-hundred and ninety-six applicants were offered an opportunity to participate in an MMI interview, and 115 completed the process. Reliability of the average of the 10 stations of the MMI was assessed using generalizability theory. A candidate by station ANOVA was performed to determine the degrees of freedom, mean squares, and estimated variance. Estimated variances were entered into the formula G-coefficient = σ^2 (candidate) / σ^2 (candidate) + σ^2 (candidate * station/10). The result was an overall test generalizability of r = 0.65. No station correlated with another station greater than r = 0.37. In addition, the overall MMI scores did not correlate with any other tool used during the admission process which included personal interview r = 0.185, simulated tutorial r = 0.32, undergraduate grade r = - 0.23, and autobiographical sketch r = 0.17. The statistical method used for reliability was not described. Validity was not examined at the time of the study.

Eva et al. (2012) expanded their work and studied the predictive validity of the MMI tool. Comparisons were made between tools used in the admission process which consisted of GPA scores, an autobiographical statement, and scores on a 12-station MMI. After applicants completed the MMI, GPA and MMI results underwent a Z score transformation and were combined. The admission committee made a decision to change the admission process based on evidence the MMI improved the association between admissions data and clinical performance. Therefore, the transformation was weighted with 30% of the weight placed on GPA and 70% on the MMI. Weighting of the autobiographical statement was not mentioned in the study. There were 1,071 students were brought in for an MMI interview, and 521 (48.6%) were admitted into the program. The accepted students had significantly higher scores on both grade point average and

the MMI, a fact not highlighted by the authors. The results were as follows; 1.) Grade point average accepted (*M* = 3.85, *SD* = 0.13, 95% CI [3.83-3.86]) versus rejected (*M* = 3.78, *SD* = 0.14, 95% CI [3.76-3.79]), *t* = 5.93, $p \leq 0.001$, *d* = 0.62, and 2.) MMI accepted (*M* = 70.5, *SD* = 10.87, 95% CI [69.6-71.5]) versus rejected (*M* = 59.4, *SD* = 11.06, 95% CI [58.1-60.6]), *t* = 11.08, $p \leq 0$ .001, *d* = 0.52 (Eva et al., 2012).

According to Eva et al. (2012), after all students matriculated through medical school, performance on the Me*dical Council of Canada Qualifying Examination* (MCCQE) was compared between the students accepted into the program *(N* = 521) and those who were rejected at the university where the study was conducted, but accepted somewhere else *(N* = 550). A total of 70.1% (751/1071) of interviewees were matched to scores on the MCCQE Part I, a multiple choice and short answer computer based examination completed shortly after graduation from medical school. Only 82.9% (623/751) of the individuals with matched scores on Part I had matched scores on the MCCQE Part II. Part II is an objective structured clinical examination typically taken 16 months into residency training. It was concluded not all interviewees had completed Part II at the time the study was conducted. The matched sample included 90.6% (472/521) accepted candidates and 50.7% (279/550) rejected candidates (Eva, et al., 2012).

Univariate analysis was performed on the MCCQE scores to examine differences between interviewees admitted to the authors' university and those accepted someplace else (rejected). Candidates accepted into the program outperformed those who were rejected both on Part I (*M* = 531, *SD* = 72.1, 95% CI [524-537] vs. *M* = 515, *SD* = 66.3, 95% CI [507-522]), *F* = 8.3, *p* = 0.003, *d* = 0.24, and Part II (*M* = 563, *SD* = 73.0, 95% CI [556-570] vs. *M* = 544, *SD* = 72.5, 95% CI [534-554]), *F* = 7.2, *p* = 0.007, *d* = 0.26 of the

MCCQE. To ensure curriculum did not impact MCCQE performance, scores of those accepted and matriculated at the authors' institution were compared to those accepted but matriculated elsewhere. The accepted / matriculated students did not outperform the accepted / matriculated elsewhere students on any outcome; Part I (*M* = 524, 95% CI [515-533] versus *M* = 546, 95% CI [535-557]), p = 0.004, and Part II (*M* = 557, 95% CI [548-566] versus *M* = *582,* 95% CI [569-594]), p = 0.003. It was concluded institutional curriculum did not impact the outcomes on the MCCQE (Eva et al., 2012).

MMI continued to be used in medical school admissions for a number of years. Over the course of two years, 484 applicants into three specialized medical programs (obstetrics-gynecology, pediatrics, and internal medicine) at one Canadian university rotated through seven MMI stations. These applicants were rated on a nine-point anchored scale, although the details were not provided (Dore et al., 2010). Generalizability theory in a cross design was used to assess three types of reliability as well as overall reliability. The internal consistency or inter-item was *r* = 0.97 - 0.98, interrater for stations with two raters was *r* = 0.78 - 0.85, interstation was *r* = 0.08 – 0.26, and the overall *r* = 0.55 – 0.70. Generalizability variance components were also assessed. The candidate x item was 0.001 – 0.01, candidate by rater was 0.36 – 0.75, and candidate by station was 1.26 – 1.96. In general, reliability was low between stations with high variance, which might be expected as each question measured different constructs. It was reported each item had high reliability with low variance and the interrater reliability was acceptable although the variance had a large range.

Husband and Dowell (2013) compared the MMI with other outcome measures in a medical school in the United Kingdom to determine predictive validity. As part of the

admission process to this medical school, four pre-admission variables were used; academic scores (school grades, aptitude testing), non-academic scores (personal statements of non-academic work), UKCAT (an intelligence test used to assess a range of mental abilities identified by medical and dental schools as important), and a 10-station MMI (Husbands & Dowell, 2013). Data were collected over two years for two cohorts of students.

Pearson correlations were conducted to examine the relationships between the four pre-admission variables, the demographic variables of age and gender, and examination scores during the program. There was no adjustment for inflation of Type I errors. However, correlations were adjusted for range restrictions (r_u) to correct for underestimates when the sample did not represent the population of interest (Husbands & Dowell, 2013).

In 2009, the Year 1 participants (n = 140) in Husbands and Dowell's (2013) study were matched to scores on written examinations during semesters 1 and 2 in the program, as well as objective structured clinical examinations (OSCE) during the same time periods. Year 2 participants (n = 128) were matched to one written and one clinical examination. During 2010, Year 1 participants (n = 150) were again matched to the four examinations described above. Data were not collected on Year 2 subjects during 2010. Statistically significant correlations were found: UKCAT scores showed significant correlations only with 2009 (Year 1) semester 1 written scores $r = 0.25$, $r_u = 0.34$, $p = 0.01$; and semester 1 OSCE scores $r = 0.18$, $r_u = 0.24$, $p = 0.03$. The MMI was significantly correlated with six of 10 data collections points: 2009 (Year 1) semester 1 OSCE $r = 0.19$, $r_u = 0.24$, $p = 0.02$; semester 2 written $r = 0.26$, $r_u = 0.33$, $p = 0.01$; semester 2 OSCE $r =$

0.34, r_u = 0.43, p = 0.01; 2009 (Year 2) written r = 0.18, r_u = 0.23, p = 0.04; and OSCE r = 0.27, r_u = 0.35, p = 0.01; and 2010 (Year 1) semester 2 OSCE r = 0.35, r_u = 0.50, $p \leq$ 0.001 (Husbands & Dowell, 2013). These results suggested a small (d = 0.2) to medium (d = 0.5) effect size (Field, 2018).

Forward entry ordinary least squares multiple regressions were also performed adding the highest simple corrrelation first and subsequent correlations next. When there was only one significant predictor, stepwise regression converted to a simple linear regression. Six significant predictors were reported. For participants in Year 1 (2009), UKCAT scores explained 6% of the variance in the semester 1 written exam, R^2 = 0.06, F = 8.81, p = 0.004 (β = 0.36, p = 0.004). UKCAT ($\beta = 9.71^{-5}$, p = 0.033) and MMI scores ($\beta = 1.79^{-3}$, p = 0.034) explained 7% of the variance in the semester 1 OSCE, R^2 = 0.07, F = 4.75, p = 0.01. MMI scores ($\beta = 2.61^{-3}$, $p \leq$ 0.001) and gender (β = -0.03, p = 0.003) explained 17% of the variance in the Semester 2 OSCE, R^2 = 0.17, F = 13.78, $p \leq$ 0.001. For participants in Year 2 (2009), MMI scores (β = 0.18, p = 0.018) and gender (β = -3.86, p = 0.007) explained 9% of the variance in the written assessement, R^2 = 0.09, F = 6.12, p = 0.003, and 15% of the variance in the OSCE, R^2 = 0.15, F = 10.72, $p \leq$ 0.001, (MMI (β = 0.15, $p \leq$ 0.001), gender (β = -2.65, p = 0.001)). For participants in Year 1 (2010), MMI ($\beta = 2.00^{-3}$, $p \leq$ 0.001) and gender (β = -0.02, p = 0.021) explained 16% of the variance in OSCE scores of semester 2, R^2 = 0.16, F = 13.56, $p \leq$ 0.001. MMI was the most consistent predictor of medical school assessments (Husbands & Dowell, 2013).

In 2017, the MMI was used in a study conducted in a Korean university to examine psychometric properties of the assessment process (Kim et al., 2017). A committee developed a six station MMI based upon constructs which were found to overlap between

competencies in the school's educational goals and the American Association of Medical Schools 15 core competencies for students entering a medical program. The six constructs were basic science, problem-solving, critical thinking, ethical decision-making, interpersonal skills, and self-regulation. A total of 164 candidates completed the study. Using variance component method, the G-coefficient of MMI scores was reported at 0.88 using the formula G-coefficient = σ^2 (candidate) / σ^2 (candidate) + σ^2 (candidate * station/6). Interrater reliability was assessed for only two of the six stations and ranged from r = 0.58 – 0.75. Independent t-tests and one-way ANOVAs were used to compare the candidates MMI scores across several variables. Scores were not significantly different based upon gender, t = 0.35, p = 0.7; undergraduate background F = 2.15, p = 0.08; or age r = 0.01, p = 0.97. Degrees of freedom were not reported. Using Pearson correlation analysis, MMI scores were not found to be associated with undergraduate GPA or scores on the Medical Education Eligibility Test (MEET). It was concluded the MMI was not biased based upon candidates' backgrounds and it assessed attributes which differed from traditional measures of cognitive abilities (Kim et al., 2017).

Jerant et al. (2017) conducted a study based upon data from five public medical schools in California. Three schools used traditional interviews and two used the MMI assessment process. Data from 4993 applicants, representing 7,516 interviews, were used for analysis. Inter-rater (inter-interviewer) or within institution reliability was calculated using Cronbach's α. It was found the correlations were generally lower between schools using the traditional interview, α = 0.13, 0.40, and 0.61, than between MMI schools, α = -.60 and 0.68. Pairwise Pearson correlations compared scores from applicants who applied at more than one school. The total interview score was converted

to a z-score (M = 1, SD = 1) to allow comparisons between schools. It was found the correlations varied considerably between schools, r = 0.18 – 0.48, with highest correlation between the schools using MMIs, r = 0.48. Finally, intraclass correlation coefficients (ICCs) were conducted comparing the MMI schools with those using traditional interviews.

All applicants who interviewed at schools were traditional interviews (TI) were used were in the TI-ICC analysis and those who were interviewed at either MMI school were in the MMI-ICC analysis. The formula for the ICC was the ratio of the variance component associated with the random effect (applicant) divided by the total variance (Jerant et al., 2017). ICC results were higher for MMI schools (0.45, 95% CI [0.40-0.54]) than interview schools (0.30, 95% CI [0.24-0.37]). ICC scores were adjusted to applicant characteristics, application year, and number and temporal sequencing of interview with similar results; MMI schools (0.47, 95% CI [0.41-0.54]) and interview schools (0.27, 95% CI [0.20 – 0.35)). It was concluded the MMI resulted in higher within and between-school reliabilities. Furthermore, applicant socio-demographic had little impact on the reliability of the instruments. A difference in internal consistency for the two MMI schools (0.60 versus 0.68) was noted. The school with the lower score had only seven stations while the school with the higher score had 10 stations. These results indicated a choice to include more stations when designing an MMI assessment.

Over time, use of the MMI expanded into the admission processes of other medical schools and health care professions and a number of qualitative studies were done. Grice (2014) reported using the MMI in the admission process to an occupational therapy program. One-hundred and six of 140 applicants were interviewed in a six station MMI. It

was concluded 98% of applicants found the process satisfactory, with 78% reporting they were ‘very satisfied” (Grice, 2014). Faculty reported the MMI was fun and allowed them to meet every applicant. None of the results assessed the psychometric properties of the assessment process.

Oyler et al. (2014) reported similar results when using a four-station MMI with students applying for entry in a pharmacy school. Thirty-seven candidates were interviewed and provided feedback. The MMI allowed them to convey their thoughts, but they did not feel this was more effective than a traditional interview. In contrast, interviewers reported feeling the MMI was more effective at assessing thoughts, skills, and processes than the traditional interview (Oyler et al., 2014). Again, no psychometric analysis was conducted.

A qualitative study was performed in one physical therapy program in Canada looking at the experiences of 18 interviewers (6 faculty, 6 clinicians, and 6 second-year students) during the MMI process (van der Spuy et al., 2016). Data were collected using semi-structured one-on-one interviews conducted in person or over the phone by two investigators. All participants acknowledged interpersonal characteristics were important to collect and the MMI helped distinguish indiviudals who were not suitable for the physical therapy profession. In addition, participants felt criterion-based scoring (using a 10-point scale range from 1= unsatisfactory to 10 = exceptional) was a more fair and objective way to score candidates than a rank-based system where each candidate was assigned a single score relative to the other candidates in the same circuit.

Over time, several systematic reviews were conducted for the MMI. Pau et al. (2013) examined CINAHL and Medline databased and found 30 studies which were

related to education and MMI. Of these studies, 24 were cross-sectional studies, three were cross-sectional with qualitative designs, and three were longitudinal in nature. Reliabilty was reported in 18 studies and found to range from moderate to high, α = 0.69 - 0.98, *G* = 0.55 – 0.72. Pau et al. (2013) indicated a need to examine reliability for groups of stations which assess the same attributes, or between group of stations examining different applicant characteristics. The MMI did not correlate with traditional assessments used in medical school admissions such as GRE and GPA, which may have indicated the MMI did examine non-cognitive attributes of applicants. The MMI was reported to have statistically predictive validity for performance at future examinations. However, no test results were reported to support this conclusion (Pau et al., 2013).

Rees et al. (2016) conducted a systematic review with results which were slightly more critical of the MMI. A total of 4,338 citations were screened by two reviewers using a Likert scale for appropriateness of design, study implementation, and data analysis. Forty-one studies were incuded in the paper. It was concluded MMIs had reasonable reliablity, α = 0.6 – 0.87. However, greater reliability was observed when the number of stations increased. Greater evidence was needed for both content and predictive validity. It was reported the MMI appeared to disadvantage rural applicants, and the possibility of an urban bias should be explored. It was acknowledged there was a need for both longitudinal studies and multi-institutional studies (Rees et al., 2016).

Non-cognitive Variables Assessed with CMSENS and CASPeR

Dore et al (2009) expanded on the previous work of the MMI by developing a new tool called the Computer-Based Multiple Sample Evaluation of Non-cognitive Skills (CMSENS). The rationale for the new tool was although the MMI had reported

correlations with clinical and non-cognitive performance of applicants in the range of *r* = 0.35—0.57, the tool could only be used with applicants who interviewed on campus. This meant reliance on typical cognitive measures determined who was invited for additional screening.

The CMSENS was designed to include eight case vignettes which were 60-90 seconds in duration and four self-descriptive questions which were similar to traditional interview questions (e.g. “What makes your heart sing?”). Each video and self-descriptive scenario had three related questions an applicant would answer. The videos were designed by experts to focus on nonmedical expert qualities (collaboration, communication, professionalism, and confidentiality). One hundred and ten applicants participated in the study consisting of 82 candidates who had been invited to interview at the university where the study was conducted, and 28 pseudo candidates who had applied to the university, but were turned down for interview. Seventy-eight participants verbally recorded responses to the questions, and the remaining 32 participants typed responses (Dore et al., 2009).

The overall reliability of the entire CMSENS tool was reported upon, although the specific type of analysis was not described. Results were 0.86 for the audio CMSENS and 0.72 for the typewritten version. Using Pearson correlation, interrater reliability was *r* = 0.82 for audio and *r* = 0.81 for typewritten versions. The typewritten CMSENS correlated with the MMI at *r* = 0.51. The audio CMSENS correlated with the MMI at only *r* = 0.15. Furthermore, scoring of the audio version took 20 minutes per scenario compared to two minutes per scenario on the typed version (Dore et al., 2009).

Because the audio version took longer and there was potential bias listening to the

recorded responses, Dore et al. (2009) continued a second part of their study using only typed responses from candidates. As before, eight 60-second vidio vignettes were included along with six self-descriptive scenarios. Candidates responded to three related questions for each scenario. Two independent raters assessed responses to each scenario using a nine-point Likert scale which ranged from “Unacceptable” to “Superior”. It was reported the overall test generalizability (statistical analysis not described) was 0.83 for CMSENS total score ($CMSENS_T$), 0.75 for the video scenarios ($CMSENS_V$), and 0.69 for descriptive scenarios ($CMSENS_D$). Pearson correlations were conducted for each type of CMSENS and the MCAT and MMI. Correlations with the MCAT were r = 0.28 $CMSENS_T$, r = 0.28 $CMSENS_V$, and r = 0.18 $CMSENS_D$. Correlations with the MMI were r = 0.46 $CMSENS_T$, r = 0.51 $CMSENS_V$ and r = 0.33 $CMSENS_D$. It was concluded the CMSENS was more closely correlated to the MMI than MCAT, and therefore more likely related to noncogntive attributes of participants (Dore et al., 2009).

Because the MMI could not be broadly administered, and the CMSNES had only moderate correlation to the MMI, Dore et al. (2017) continued further refinement of a computerized tool. The Computer-Based Assessment for Examining Personal Characteristic (CASPeR) was developed. This tool contained 12 scenarios; four written behavioral scenarios and eight video-based scenarios called situational judgment tests. After reviewing a scenario, each candidate had five minutes to respond to three open-ended questions. It was believed the open-ended responses allowed a candidate to provide answers based upon the unique diversity and experiences each candidate experienced.

In 2012, 109 participants who had taken the CASPeR between 2007 and 2008 and

were selected and completed medical training programs across Canada, were invited to participate in the study. Of those participants, 63 had completed Part I of the medical exam (multiple choice and clinical decision making) and 53 had completed Part II of the exam (14 station Objective Structured Clinical Examination (OSCE)).

Bivariate correlations were conducted using a dis-attenuation correction as follows: $Rxy = rxy / \sqrt{(rxxx\ ryy)}$. Based on the results, the four written behavioral scenarios of CASPeR did not significantly correlate with the three professional domains of the medical licensing examination (MCCQE); 1.) Part I - CLEO (communication, legal, ethical, and organization, 2.) Part I - PHLEO (public health, legal, ethical and organizational, or 3.) Part II – CLEO. In contrast, the eight situational judgments tests were significantly correlated with the professional domains of the MCCQE at a moderate level (Part I CLEO, $r = 0.30$, $p = .038$; Part I PHLEO, $r = 0.036$, $p = .014$; Part II CLEO, $r = 0.50$, $p = .025$) (Dore et al., 2017). Neither the situational judgment tests nor written behavioral scenarios were significantly correlated to any cognitive portion of the MCCQE (medicine, surgery, psychiatry, pediatrics, obstetrics/gynecology, or family medicine). This result was anticipated as the scenarios were designed to test non-cognitive attributes of a candidate. CASPeR had a stronger correlation on Part II CLEO than the MMI. Part II of the medical examination was entirely based on objective structured clinical examinations and contained no multiple choice questions, which were thought to assess cognitive attributes.

Development of CANA-HP

At the present time, CASPeR is a proprietary owned assessment tool, which is being used by applicants to medical school, and physician assistant and physical therapy programs. The tool has not been piloted with other health care professions, and is not

available for psychometric testing with these populations. The CASPeR was developed to include eight situational judgments tests (SJT) and four behavioral scenarios. The SJTs had better correlation to the MMI than the four descriptive scenarios (Dore et al., 2009), and will therefore form the basis for a new assessment tool.

Patterson et al. (2016), in an overview of best evidence, described the SJT as a measurement methodology where a candidate is given a situation which might be encountered during a professional role, and the candidate selects a response from a pre-determined list of possible options which might include multiple choice, ranking, or single best answer (Figure 1). Each SJT response was scored by comparing candidates' responses to a pre-determined scoring key. In the overview of the evidence for SJTs, Patterson et al. (2016) reported in medical education SJTs had internal consistency, $\propto$ = 0.43 - 0.94, parallel reliability, r = 0.66 - 0.76, criterion related validity, r = 0.25 - 0.47, and greater predictive validity at the lower end of performers (Patterson, Zibarras, & Ashworth, 2016). Traditional SJTs were reported to be cost-effective and efficient to determine non-cognitive attributes of applicants.

Because CASPeR is a proprietary tool, a literature review was conducted to find other non-cognitive tools used for holistic admission. In 1976, Sedlack and Brooks identified eight non-cognitive dimensions of students which were thought to be important to the success of minority students. These eight dimensions included academic positive self-concept, realistic self-appraisal, support of academic plans, leadership, long range goals, ability to establish community ties, understanding of racism, academic familiarity (Sedlacek & Brooks, 1976).

Tracey and Sedlacek (1984) developed the *Non-Cognitive Questionnaire*, which

was designed to measure both creative and practical abilities of individuals. It incorporated the eight dimensions previously identified. The Non-Cognitive Questionnaire was reported to have internal consistency between 0.37 - 0.82 for Caucasian students and 0.49 – 0.84 for African American students (Tracey & Sedlacek, 1989).

Subsequently, the decision was made to further refine the first instrument to enable the subscales to more accurately reflect the desired constructs (Tracey & Sedlacek, 1989). The original questionnaire contained only 1-3 items per construct. The Non-Cognitive Questionnaire-Revised (NCQ-R) contained 38 items related to the same eight non-constructs, however, each construct was now was represented by 3-7 items. The subscale structure of the NCQ-R was examined using Confirmatory Factor Analysis on the item covariance matrix using the LISREL VI package. Because the factor structure had not yet been determined, initial estimates of loading were conducted using the minority population (black sample). Because one factor loading may not be representative of other samples, the black sample was further split into two subsamples; the first for parameter estimation (n = 101) and the second to test generalizability of the results (n = 97). Finally, the parameters estimates across race was examined by including a random sample of white students (n = 222) (Tracey & Sedlacek, 1989).

Prior to conducting the factor analysis, the internal consistency of the eight dimensions was estimated using Cronbach's alpha. Results were Black sample-1, α = 0.55 – 0.84, Black sample-2, α = .0.49 – 0.83, and White sample, α = 0.37 = 0.70. The White and Black-2 subsets had lower reliability on academic self-concept and academic self-plans. However, these constructs had the fewest number of items in them which was reported as a possible contribution to the variability. Internal consistency was also lower

for the White subset on racism, α = 0.37 (Tracey & Sedlacek, 1989). It was concluded the validity of this test among Whites may be questionable. There was a fair amount of overlap among the eight constructs, especially in racism and realistic self-appraisal. It was hypothesized these constructs may be difficult to define or the constructs may be important in the remaining six attributes.

Goodness of fit was determined using three indices; goodness-of-fit index (GFI), the root mean square residual (RMR), and the Tucker and Lewis index (TL1), which is a reliability coefficient for maximum likelihood factor analysis (Anderson & Gerbing, 1984). Results were reported for Black sample-1 as follows: *GFI* = 0.83, *RMS* = 0.42, and *TL1* = 0.85. The invariance of the model was examined with both the second black sample and the white sample. For the Black-1 versus Black-2 subsamples the following goodness-of—fit indices were obtained; *GF1* = 0.77, *RMR* = 0.71, *TL1* = 0.72. For the Black-1 versus White sample the results were as follows; *GF1* = 0.84, *RMR* = 0.45, and *TL1* = 0.73. It was concluded the fit of the two subsamples was generally adequate (Tracey & Sedlacek, 1989). One of the problems with confirmatory factor analysis, when using three or more indicators in a factor, is the minimum required sample size, which was 150, to obtain solutions which are proper and convergent (Anderson & Gerbing, 1984). Two of the subsamples used for confirmatory factory analysis Black-1 (n = 101) and Black-2 (n = 97) contained fewer than 150 participants.

Sedlacek (2017) slightly changed the titles of the eight constructs as follows: positive self-concept, realistic self-appraisal, understands and knows how to navigate the system and racism, prefers long-range goals to short-term or immediate needs, availability of a strong support system, successful leadership skills, demonstrated

community service, and knowledge acquired in or about a field (nontraditional learning). These non-cognitive variables were developed to improve admission, success, and retention for under-represented students. However, the variables were not specific to professional attributes for specific health professions to be addressed in this study, nurse anesthesia, occupational therapy, physician assistant studies, and physical therapy. Therefore, additional professional attributes were sought from the literature.

In nurse anesthesia, eight professional attributes represented the non-clinical skills, attitudes, and judgments fundamental for success in the field (AANA, 2016). These attributes were identified as collaborative, culturally competent, evidence based practice, leader, professionally engaged, situationally aware, teacher, and well. In occupational therapy, seven core values were identified to serve as the basis for the profession (Kanny, 1993). These core values were altruism, equalilty, freedom, justice, dignity, truth, and prudence. Among physician assistants, six professional competencies were identifed (ARC-PA et al., 2012) and included medical knowledge, interpersonal and communication skills, patient care, professionalism, practice-based learning and improvement, and systems-based practice. For physical therapists, the necessary skills for the profession were determined to be a set of seven core values (APTA, 2010). These core values were accountability, altruism, compassion / caring, excellence, integrity, professional duty, and social responsbiity. These attributes will be used to develop a new tool called the CANA-HP.

CHAPTER 3 – METHODOLOGY

Research Question

The primary purpose of this study was to examine the psychometric properties of the *Computer-based Assessment of Non-cognitive Attributes of Health Professionals* (CANA-HP). Three specific questions were delineated, focusing on different aspects of the CANA-HP:

1. What is the CANA-HP instrument reliability (internal consistency & interrater) within each station (rater) and inter-station (station)?
2. Does the CANA-HP measure attributes of non-cognitive variables as demonstrated by low construct validity scores when correlating the CANA-HP to traditional assessments reported to measure cognition (e.g. pre-admission GRE and GPA)?
3. Does analysis reveal differences between groups based upon gender, ethnicity, Pell-grant status, family history of college, or socio-economic differences?

Wayne State University is a pubic, research intensive university located in the urban community of Detroit, Michigan. The university houses fourteen schools and colleges in a variety of disciplines, including the Eugene Applebaum College of Pharmacy and Health Sciences (EACPHS). Students are admitted into EACPHS seeking education in one of twelve degree granting professional programs, most at the graduate level. The population the CANA-HP is intended to be used with includes applicants seeking admission into four health care professional programs offered at EACPHS; nurse anesthesia, occupational therapy, physician assistant, and physical therapy. For this

study, data from applicants into the occupational therapy program were examined.

Annually, admission committee members from each of these four programs review applicants and interview top candidates to fill the cohort of incoming students for the upcoming academic year. Applicants selected for an interview must have met admission criteria (which vary slightly for each program) and have a minimum GPA of 3.0. Table 2 contains information regarding applicant status from the 2017-2018 applicant pool for each of the four programs.

Table 2.

Study Population across the Four Health Profession Programs

Program	# Qualified Applicants	# Interviewed	# Accepted
Nurse Anesthesia	155	86 (55%)	24 (15%)
Occupational Therapy	88	74 (84%)	33 (38%)
Physician Assistant	350	150 (43%)	50 (14%)
Physical Therapist	237	157 (67%)	33 (14%)

Participants

Prior to recruiting applicants to serve as participants for this study, Human Subject Approval to conduct research with human subjects was obtained from the Wayne State University Institutional Review Board via an expedited review for behavioral research (IRB 19-12-1558)(Appendix A). All applicants who accepted an invitation for an admission interview into the occupational therapy program were invited to participate in the study. This yielded a convenience sample composed of voluntary participants. All of the applicants selected for interview were advised the assessment to be administered was solely for psychometric property purposes, and refusal to complete the assessment would not impact the application process. Participants were informed they could withdraw from

the study at any time. Inclusion criteria was limited to the applicants selected for in-person interviews due to a desire to ensure the applicant was the person completing the assessment. Applicants were excluded from the study if they did not sign an informed consent, if they were not selected to interview for the occupational therapy program, or if the applicant was not at least 18 years of age.

Instrument Development

The *Computer-based Assessment of Non-cognitive Attributes of Health Professionals* (CANA-HP) represents a novel methodology designed to measure specific non-cognitive attributes of applicants seeking admission into a health care profession. The CANA-HP was developed by comparing profession specific attributes of four health professions to the eight non-cognitive variables developed by Sedlacek (2017).

Table 3 represents the overlap between Sedlacek's non-cognitive variables and the professional attributes of the four health professions to be included in the larger study. A total of six non-cognitive factors were identified as applicable to all four programs; positive self-concept, realistic self-appraisal, ability to navigate systems and cultures, leadership, community service, and interpersonal skills & communication. Although communication and interpersonal skills were not part of Sedlacek's original eight variables (Sedlacek, 2017), three of the four programs highlighted this variable as critical to the profession (AANA, 2016; APTA, 2010; ARC-PA, 2012), therefore the variable was included in this study. Three non-cognitive values were not included. Delayed gratification and strong support system were not identified by any of the four professions as a core value. Knowledge of field was considered to profession dependent and, therefore, situational judgments tests applicable to four different health care professions may have

been difficult to develop. Therefore, the six non-cognitive attributes included in the CANA-HP are defined below.

1.) *Positive self-concept*: The student expresses confidence, strength of character, determination and independence.

2.) *Realistic self-appraisal*: The students has recognition and acceptance of strengths and deficits, especially academic. The student works on self-development, applies critical thinking, and recognizes a need to broaden his/her individuality.

3.) Ability to navigate system and culture: The student exhibits a realistic view of the system based upon experiences, is committed to improving the system, and takes an assertive approach to dealing with wrongs. The student is not hostile to society.

4.) Leadership: The student demonstrates leadership in any area of background (church, sports, non-educational groups).

5.) Community service: The student participates in and is involved in the community and cares about the welfare of others.

6.) Communication and interpersonal: The student demonstrates effective interpersonal and communication skills. The student is able to identify a sense of caring about another individual's welfare.

Table 3.

Comparison of Non-cognitive Attributes with Profession Specific Attributes

Non-Cognitive Variables (Sedlacek, 2017)	Nurse Anesthesia (AANA, 2016)	Occupational Therapy (Kanny, 1993)	Physician Assistant (ARC-PA, 2012)	Physical Therapist (APTA, 2010)

Positive self-concept	Well	Freedom	Professionalism	Excellence
Realistic self-appraisal	Situationally aware	Truth	Practice-based learning & improvement	Accountability
Ability to navigate systems & culture	Culturally competent	Dignity / Equality / Justice	Systems-based practice	Social responsibility / Integrity
Delayed gratification				
Strong support system				
Leadership	Leader / Teacher	Prudence	Patient Care	Professional Duty
Community service	Professionally engaged	Altruism		Altruism
Knowledge of field	Evidence based practice		Medical Knowledge	
Communication & interpersonal*	Collaborative		Interpersonal & communication skills	Compassion / Caring

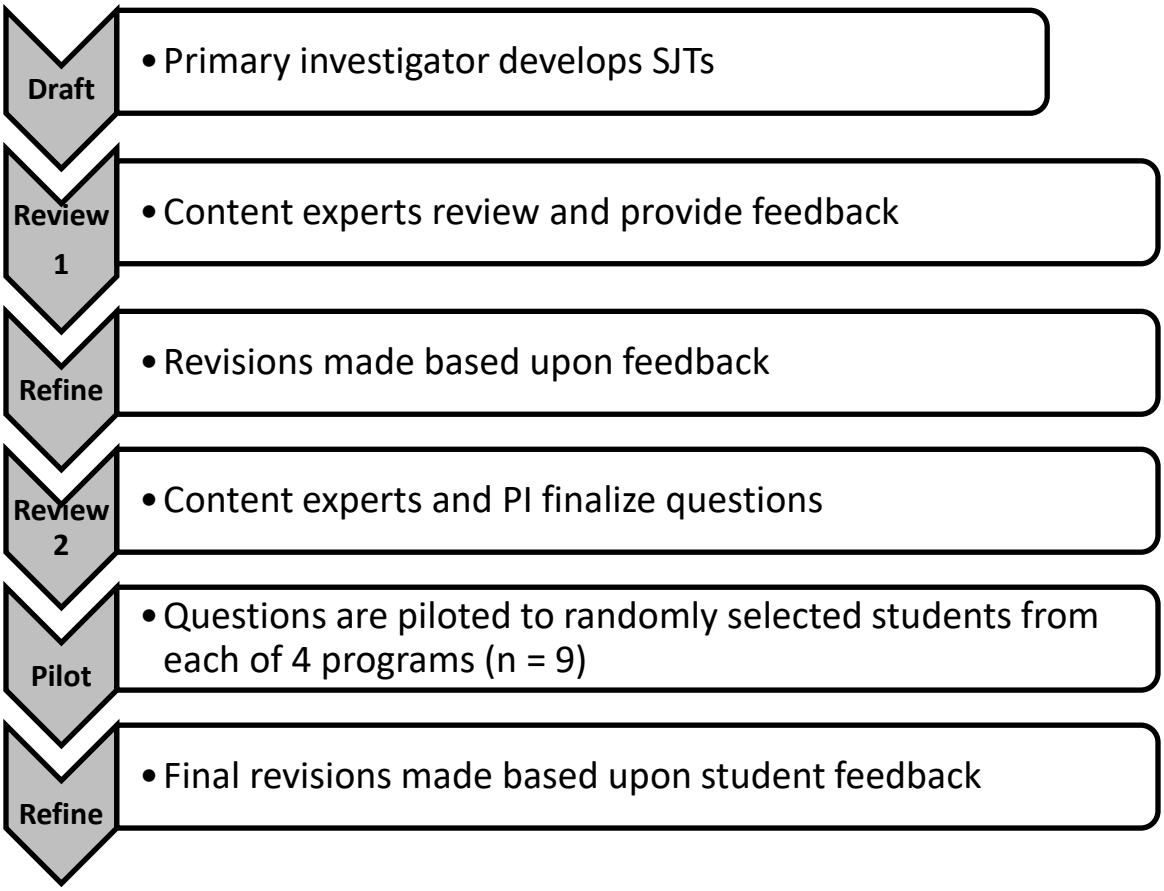
*Not part of Sedlacek's original eight non-cognitive attributes

The CANA-HP was designed as a computer assessment consisting of 12 stations, each containing a situational judgment test (Appendix B). Six of the stations contained situational judgment tests with open-ended questions and six contained a situational judgment test in a traditional format (multiple choice, ranking best answer). A separate question was developed for each of six non-cognitive variables; positive self-concept, realistic self-appraisal, able to navigate systems & cultures, leadership, and community

service, for both types of stations (open-ended and traditional). The non-cognitive attribute of communication and interpersonal skills were woven into the six open-ended stations. Outlined in Figure x was the plan for development of the situational judgment tests.

Figure 1.

Development of Situational Judgment Tests (SJTs)



Reliability and Validity

According to Fraenkel et al. (2016), reliability refers to the consistency of scores obtained from one individual administration to another administration, or from one set of items to a different set of items. Reliability has three general forms; test-retest, equivalent-forms (also known as alternative or parallel), and internal-consistency methods (Fraenkel,

2016). When describing the results of reliability, statements of the results should be accompanied by an explanation of the type of reliability performed, how the results were calculated, and the conditions under which each result was obtained (Sawilowsky, 2000, p. 159). Internal consistency of the CANA-HP was obtained for both interrater and station.

Validity is ‘the degree to which a test measures what it purports to measure, and relates to the use of the test as opposed to the test itself” (Sawilowsky, 2000, p. 166). According to Fraenkel et al. (2016), there are three primary types of validity: content, criterion (both predictive and concurrent), and construct. Predictive validity pertains to how well scores on one instrument will correlate with scores on a different criterion variable at a future time. Concurrent validity, on the other hand, compares and scores on an instrument to a criterion variable at the same point in time. Construct validity refers to how well a construct (such as self-esteem) actually matches a person’s ability in the construct (degree of self-esteem a person possesses). There are several methods to examine construct validity, two of which include exploratory factor analysis and confirmatory factor analysis. Construct validity evidence for the CANA-HP was obtained by comparing scores on the traditional and open-ended situationa judgments test with GRE and GPA scores. The hyothesis was there would be no correlation between these items.

Procedures

Applicants into the occupational therapy program were sent an electronic invitation to participate in the study, after they have been invited for an admission interview by the Chair of Admissions for the program. The invitation to participate in the study also included an electronic copy of the informed consent for the applicant to review, and

instructions on how to participate on the day of the interview.

On the day of the interview, the primary investigator or research assistant met with the applicants in a computer lab. (The number of students varied depending on how many applicants were brought in by the program for interviews at the same time). The primary investigator or research assistant described the study, provided a short overview of the informed consent, and answered any questions the applicants had. All participants were given the password to the survey.

The first question asked for informed consent. If consent was confirmed, the applicant was directed to the next page, which included the following demographic information; unique identifier, age, sex, ethnicity, GPA, GRE score, first-generation student status, socioeconomic status, Pell-grant status, experience with disadvantaged populations, and geographic location of current living situation (urban, rural, etc.). The last demographic question asked the participant if the research team could access the admission application to retrieve verified GRE and GPA scores. If the applicant selected yes, he/she was asked to provide a unique identifier on the survey as well as write their name and identifier on a 3x5 card which was given to a member of the research team. On completion of the demographic information, or if an applicant did not agree to participate in the study, the applicant's questionnaire moved directly into the 12 situational judgment test stations. Each situational judgment test was timed with no more than 15 minutes allocated in the traditional stations (multiple choice) and no more than 45 minutes for the six open-ended stations. The participant could spend no more than 70 minutes participating in the study.

Each participant earned a total score ranging from –6 to 9 on each of the six

traditional stations (multiple choice), as well as a total score from 7 to 35 for each of the six open-ended stations. The traditional stations was scored using a grading system where correct answers were worth 3 points (3 total), neutral answers were worth 0 points, and the remaining answers had increasing negative value (-1, -2, and -3). The traditional station results were self-graded by computer software within the Qualtrics program. The open-ended stations were scored on seven, 5-point Likert scale rubrics (range of 1-5) by reviewers, who consisted of one research assistant and the primary investigator. Each applicant was scored by the two reviewers. All reviewers were trained to score all stations using rubrics specific for the station. Training was completed by having each reviewer score data from the pilot study participants until agreement was achieved between the two reviewers. All identifiable participant information was removed prior to scoring.

Data Analysis

All data were analyzed using SPSS v. 25.0 (IBM, 2018) or Iteman v. 4.3 (ASC, 2013). Descriptive statistics of the sample population were determined for each program (mean age, gender, socio-economic statues, etc.), as well as for the each of the twelve stations of the CANA-HP (mean, standard deviation).

Cronbach's coefficient alpha, a measure of internal consistency, was conducted on the twelve situational judgment tests, using alpha values $\alpha \geq 0.7$ as evidence of reliability (Fraenkel, 2016). To determine interrater reliability, interclass correlation coefficients (ICC) estimates and their 95% confidence intervals were calculated based on average rating (k = 2), consistency-agreement, 2-way way random-effects model. Values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9 and greater than 0.90 were indicative of poor, moderate, good, and excellent reliability respectively (Koo & Li, 2016).

Pearson's correlation coefficients were calculated between CANA-HP scores and GRE and GPA scores at the time of program admission. To control for Type I error, Bonferroni corrections were applied to results of the multiple comparisons. The hypothesis was the correlation would be low between these three items because the CANA-HP measures cognitive variables and GRE and GPA are cognitive measures.

The CANA-HP used partial scoring for all stations, and, therefore, the scales were considered polytomous in nature. In order to run item difficulty and item discrimination, the scores for both the traditional and open-ended stations were adjusted. Iteman software limits each variable to a maximum of 15 options (possible scores), and does not recognize negative values as a plausible outcome. Therefore, for the six traditional stations, each participant was initially given a total score ranging from –6 to 9. The scores were adjusted so each negative score was converted to a zero. All other scores remained the same. For the open-ended stations, each construct was rated on 7 characteristics using a Likert scale (scores which ranged from 1-5). The seven characteristics were totaled, for a final station score ranging from 7-35. Because of the 15 option limit, each final station score was divided by 7 (the number of characteristics) to give an average score for the station. The average scores used in item difficulty and discrimination analysis, therefore, ranged from 1-5.

Item difficulty was analyzed using mean average (*P*), and test discrimination was conducted with Pearson point-biserial correlation (*Rpbis*). The *P* value was the average of item responses converted to numeric values across examinees. A good rating scale was considered to have a mean close to 50% of the maximum score for the item (Guyer & Thompson, 2013). The *Rpbis* value ranged from -0 to 1.0 with a minimal acceptable

range starting between 0.10 – 0.20 and the maximum range rarely above 0.50. A negative point-biserial indicated a very poor item, and a score of 0.0 indicated no differentiation (Guyer & Thompson, 2013).

Fisher's exact tests, with Bonferroni adjustments for all *p* values, were conducted to determine if the CANA-HP scenarios were biased for minorities, individuals who had received Pell-grants, individuals of differing socio-economic status, or individuals who were the first generation to attend college. Due to the small sample size, all categories were collapsed to increase the number of individuals in each category. Binary categories were created for race (minority or Caucasian), Pell-Grant status (recipient or non-recipient), and family attending college (first generation or not first generation). Income and the three scores on the situational judgment tests (final written, final open-ended, and total overall score) were broken down into quartiles. The hypothesis was there would be no statistically significant difference in scores between any of the groups.

CHAPTER 4 – RESULTS

The purpose of this study was to develop a novel methodology which could effectively screen non-cognitive attributes of applicants seeking admission into one of four health care professions: (1) nurse anesthesia, (2) occupational therapy, (3) physician assistant, and (4) physical therapy. Only students who applied for admission into the occupational therapy program at the university where the study was conducted were included in the initial study.

Participants

There were N = 38 applicants interviewed in February, 2020, as part of the application process for the occupational therapy program. Thirty-seven (97.4%) of those applicants agreed (through electronic consent) to participate in the study. Demographics for these participants are compiled in Table 4. They were primarily female (86.5%), Caucasian (73%), with a mean age of 23.0 (± 3.76). All applicants had attended some college and the majority had at least one immediate family member (78.4%) who also had attended college. (Immediate family members included any one of the following individuals; grandparent, parent, aunt/uncle, or sibling.) The participants had an average undergraduate GPA of 3.51 (± 0.34), with a pre-requisite GPA's in science of 3.46 (± 0.36) and non-science of 3.61 (± 0.31).

Only 8.1% of the participants reported they were not working (unemployed) at the time of the survey. The majority worked in unskilled professional labor (64.9%) such as employment as an occupational therapy technician. The self-reported, annual household income varied considerable among the participants with a mean of $84,813 (range $15,000 - $200,000). Overall, 25% of the participants been awarded a Pell-Grant.

Although the university where the study was conducted was located in an urban setting, only 10.8% of participants lived in an urban area. The majority lived in the suburbs (83.8%).

Table 4.

Demographics

Variable	Category	Frequency	Percent
Gender	Female	32	86.5%
	Male	5	13.5%
Race/Ethnicity	Black	3	8.1%
	Hispanic	2	5.4%
	Multi-racial	1	10.8%
	Middle-eastern	4	2.7%
	White / Caucasian	27	73.0%
Employment	Unemployed	3	8.1%
	Unskilled manual	2	5.4%
	Unskilled professional	24	64.9%
	Skilled manual	1	2.7%
	Professional	7	18.9%
Current living situation	Rural	2	5.4%
	Suburban	31	83.8%
	Urban	4	10.8%

Variable	N	Mean	SD	Range
Age	37	23.00 yrs.	(+ 3.76)	20-43 yrs.
Income	30	$84,313	(+ $51,586)	$15,000 – $200,000
Science GPA	37	3.46	(+ 0.36)	2.68 - 4.0
Non-science GPA	37	3.61	(+ 0.31)	2.91 - 4.0
Undergraduate GPA	37	3.51	(+ 0.34)	2.55 - 4.0
Verbal GRE	11	145.4	(+ 5.0)	136 - 151
Quantitative GRE	11	145.6	(+ 6.0)	132 - 154
Analytic GRE	9	3.7	(+ 0.8)	2.0 - 4.5

Situational Judgment Test Individual Results

The main descriptive statistics of the 12 situational judgments tests are presented in Table 5. The first six scores (#1-6) represent findings from the traditional format stations (multiple choice) which had a range of -6 to 9. The second six scores (#7-12) represent findings from the open-ended stations which had a range from 7-35. For the traditional format, highest mean scores were obtained for realistic self-appraisal (*M* = 6.97) and leadership (*M* = 7.14). For the open-ended stations, the highest mean scores were obtained in positive self-concept (*M* = 28.78) and navigating systems / culture (*M* = 29.45).

The highest standard deviations for the traditional stations were found for navigate systems / culture (*M* = 3.27) and community service (*M* = 3.02), and the lowest standard deviations were found for realistic self-appraisal (*M* = 2.05) and communication & interpersonal (*M* = 1.94). For the open-ended stations, the highest standard deviations were found for navigate systems / culture (*M* = 5.89), and leadership (*M* = 5.30). The lowest standard deviations were found for positive self-concept (*M* = 2.38) and community service (*M* = 3.18).

Table 5.

Descriptive Statistics for 12 Situational Judgment Tests

Question Number	Type	Construct	Mean	SD	Min	Max
1	Traditional	Positive self-concept	4.76	2.66	-2	9
2	Traditional	Realistic self-appraisal	**6.97**	2.05	-1	9
3	Traditional	Navigate systems / culture	4.92	**3.27**	-2	9
4	Traditional	Leadership	**7.14**	2.31	2	9
5	Traditional	Community Service	2.49	**3.02**	-1	9
6	Traditional	Communication & interpersonal	4.43	1.94	2	9
7	Open-ended	Positive self-concept	**28.78**	2.38	23.0	33.5
8	Open-ended	Realistic self-appraisal	27.89	4.39	12.5	34.0
9	Open-ended	Navigate systems / culture	**29.45**	**5.89**	14.0	35.0

10	Open-ended	Leadership	24.85	**5.30**	14.0	33.5
11	Open-ended	Community Service	22.32	3.18	13.0	27.5
12	Open-ended	Communication & Interpersonal	21.15	4.00	14.0	27.5

Reliability of Stations

Reliability for the CANA-HP stations were assessed using Cronbach's alpha. It was $\alpha = 0.492$, which is low. Although rules of thumb abound, this magnitude did not meet even a modest criterion for evidence of reliability set at $\alpha \geq 0.7$ by Fraenkel (2016). Results for each individual item are shown in Table 6. Four of the traditional stations, #2, #4, #5, and #6 had negative corrected item-total correlation which resulted in a higher Cronbach's alpha value if these items were deleted. For this reason, the decision was made to conduct separate Cronbach's alpha analyses for both the traditional and open-ended stations.

Table 6.

Cronbach's Alpha Results for 12 Situational Judgment Tests

Number	Type	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1	Traditional	180.392	267.974	.076	.149	.494
2	Traditional	178.176	285.322	-.113	.381	.520
3	Traditional	180.230	260.605	.099	.153	.492
4	Traditional	178.014	291.687	-.194	.389	.537
5	Traditional	182.662	284.709	-.119	.461	.538
6	Traditional	180.716	292.799	-.224	.159	.533
7	Open-ended	156.365	248.328	.369	.277	.440
8	Open-ended	157.257	212.036	.394	.356	.398
9	Open-ended	155.703	163.270	.556	.563	.292
10	Open-ended	160.297	188.881	.435	.476	.367
11	Open-ended	162.824	235.864	.366	.456	.426
12	Open-ended	164.000	240.069	.207	.287	.465

Cronbach's alpha for the six traditional stations was α = 0.091 and results are compiled in Table 7. All six stations had minimal variation in Cronbach's alpha values (α = .064 – 0.171) if the item were deleted, and all values were in the low range. Because all items were coded using the same system, realistic self-appraisal was not coded incorrectly despite having a negative value. Thus, the result for realistic self-appraisal may not have had high covariance.

Table 7.

Cronbach's Alpha Values for Traditional Stations

Construct	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Positive self-concept	31.14	17.176	.029	.030	.081
Realistic self-appraisal	30.24	15.189	.230	.099	-.115[a]
Navigate systems / culture	30.97	16.527	.045	.041	.064
Leadership	29.84	18.473	-.063	.213	.171
Community service	32.51	16.757	.023	.186	.088
Communication / interpersonal	32.19	18.324	-.024	.043	.129

a. The value is negative due to a negative average covariance among items.

However, Cronbach's alpha for the open-ended stations was found to be α = 0.706, which is indicative of minimally adequate reliability. Cronbach's alpha values for each individual station can be found in Table 8. The six stations had minimal variation if deleted, which ranged from α = 0.582 to 0.717. In general, all six stations appeared to strengthen the overall reliability.

Table 8.

Cronbach's Alpha Values for Open-ended Stations

Construct	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted

Positive self-concept	125.662	246.292	.348	.198	.698
Realistic self-appraisal	126.554	198.178	.489	.289	.650
Navigate systems / culture	125.000	149.139	.654	.493	.582
Leadership	129.595	176.553	.510	.288	.643
Community service	132.122	225.020	.449	.279	.670
Communication / interpersonal	133.297	231.006	.255	.086	.717

Because the number of items for both the traditional and open-ended stations was small, a Spearman Brown correction was computed. Results are compiled in Table 9. To achieve a minimally adequate reliability for all stations, the number of items would need to be tripled (α = 0.744). However, even with triple questions the new alpha level would still remain low (α = 0.231) for the traditional stations. To achieve minimally adequate reliability for these stations, 144 questions would need to be created (α = 0.231).

Table 9.

Spearman Brown Correction for All Stations

Stations	Original Cronbach's alpha	Original number of items	New number of items	Spearman Brown Correction
	Twice the number of items			
Total Overall Score	0.492	12	24	0.660
Traditional Stations	0.091	6	12	0.167
Written Stations	0.706	6	12	0.828
	Triple the number of items			
Total Overall Score	0.492	12	36	0.744
Traditional Stations	0.091	6	18	0.231
Written Stations	0.706	6	18	0.878
	Number of items to achieve minimum α = 0.70			
Total Overall Score	0.492	12	288	0.959
Traditional Stations	0.091	6	144	0.706
Written Stations	0.706	6	144	0.983

Interrater Reliability

Reliability of the raters was assessed using ICC estimates and their 95% confidence intervals as previously described. Table 10 shows the ICC estimates for each of the six open-ended situational judgment tests. The traditional stations were multiple choice in nature and scored by the survey instrument, therefore, interrater reliability for the traditional stations is not reported. The interrater reliability for the two raters in the study ranged from ‘moderate’ for positive self-concept (0.67) and realistic self-appraisal (0.67) to ‘excellent’ for navigate systems / culture (0.91). The 95% confidence interval results, on the other hand, lead to the interpretation of ‘poor’ to ‘good’ reliability for positive self-concept and realistic self-appraisal, ‘moderate’ to ‘excellent’ reliability for community service and communication / interpersonal skills, and ‘good’ to ‘excellent’ reliability for navigate systems / culture and leadership.

Table 10.

Intraclass Correlations Coefficients for Average Measures for the Open-ended Stations

		95% Confidence Interval		F Test with True Value 0			
Construct	Intraclass Correlation[b]	Lower Bound	Upper Bound	Value	df1	df2	Sig
Positive self-concept	.672	.364	.831	3.052	36	36	.001
Realistic self-appraisal	.674	.367	.832	3.068	36	36	.001
Navigate systems / culture	.908	.822	.953	10.892	36	36	.000
Leadership	.891	.788	.944	9.168	36	36	.000
Community service	.827	.665	.911	5.789	36	36	.000
Communication / interpersonal	.817	.645	.906	5.468	36	36	.000

Note: Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

On average, the six open-ended stations took a total of 6 minutes and 29 seconds to grade. The longest station took an average of 1 minute and 22 seconds to grade, while the shortest station took an average of 50 seconds (see Table 11).

Table 11.

Average Time for Scoring the Six Open-ended Stations

Construct	Average Time	SD (in seconds)	Maximum Time	Minimum Time
Positive self-concept	**0:01:22**	0:00:24	0:02:25	0:00:51
Realistic self-appraisal	0:01:01	0:00:16	0:02:06	0:00:31
Navigate systems / culture	0:00:57	0:00:16	0:01:52	0:00:33
Leadership	0:01:13	0:00:17	0:01:49	0:00:34
Community service	0:01:06	0:00:18	0:01:56	0:00:35
Communication / interpersonal	**0:00:50**	0:00:15	0:01:46	0:00:30

*****Note: Time is written in format hours:minutes:seconds

Construct Validity

Three types of GPA (undergraduate, science, non-science) were correlated with the final score on the six traditional stations, the final score on the six open-ended stations and the total overall score for all twelve stations. Descriptive statistics for each of the GPAs and the station totals are shown in Table 12.

Table 12.

Descriptive Statistics for GPA and Stations Scores

			Bootstrap[a]			
					95% Confidence Interval	
Variable		Statistic	Bias	Std. Error	Lower	Upper
Science GPA	Mean	3.4624	.0009	.0579	3.3500	3.5743
	Std. Deviation	.35523	-.00607	.03278	.28532	.41349
	N	37	0	0	37	37
Non-science GPA	Mean	3.6089	.0005	.0494	3.5100	3.7038
	Std. Deviation	.30573	-.00565	.02775	.24496	.35245
	N	37	0	0	37	37
Undergraduate GPA	Mean	3.5092	.0002	.0559	3.3951	3.6143
	Std. Deviation	.34228	-.00673	.04097	.25577	.41625
	N	37	0	0	37	37
Total Traditional	Mean	30.70	.01	1.07	28.57	32.73
	Std. Deviation	6.591	-.147	.826	4.846	8.048
	N	37	0	0	37	37
Total Open-Ended	Mean	154.446	.035	2.721	148.784	159.459
	Std. Deviation	16.6712	-.5507	2.9685	10.3982	21.7761
	N	37	0	0	37	37
Total Overall Score	Mean	185.1486	.0495	2.7303	179.6486	190.3243
	Std. Deviation	16.78370	-.47154	2.50279	11.50467	21.16609
	N	37	0	0	37	37

a. Unless otherwise noted, bootstrap results are based on 10000 bootstrap samples

Results of the Pearson correlation analysis can be seen in Table 13. All three GPA scores were significantly correlated to each other ($p \leq .001$). The final score on the traditional stations (multiple choice) was not correlated to any of the GPA scores ($p = .084$ - .699). However, non-science GPA was significantly correlated to the final score on the open-ended stations ($p = .002$) and total overall score ($p = .008$).

Table 13.

Correlations between Three Types of GPA and Total Station Scores

Variable		Science GPA	Non-science GPA	Undergraduate GPA	Total Traditional	Total Open-ended	Total Overall Score
Science GPA	Pearson Correlation	1	.678**	.657**	-.066	.269	.241
	Sig. (2-tailed)		**.000**	**.000**	.699	.107	.150
	N	37	37	37	37	37	37
Non-science GPA	Pearson Correlation	.678**	1	.867**	-.163	.496**	.429**
	Sig. (2-tailed)	**.000**		**.000**	.335	**.002**	**.008**
	N	37	37	37	37	37	37
Undergraduate GPA	Pearson Correlation	.657**	.867**	1	-.288	.326	.210
	Sig. (2-tailed)	**.000**	**.000**		.084	.049	.211
	N	37	37	37	37	37	37
Total Traditional	Pearson Correlation	-.066	-.163	-.288	1	-.181	.213
	Sig. (2-tailed)	.699	.335	.084		.285	.205
	N	37	37	37	37	37	37
Total Open-ended	Pearson Correlation	.269	.496**	.326	-.181	1	.922**
	Sig. (2-tailed)	.107	**.002**	.049	.285		**.000**
	N	37	37	37	37	37	37
Total Overall Score	Pearson Correlation	.241	.429**	.210	.213	.922**	1
	Sig. (2-tailed)	.150	**.008**	.211	.205	**.000**	
	N	37	37	37	37	37	37

** Correlation is significant with a Bonferroni adjustment (p = .05/5 = .01).

Pearson's correlation was conducted for the three types of GRE scores (verbal, quantitative, and analytic) and the three station scores (final score on the six traditional stations, final score on the six open-ended stations, and total overall score). GRE scores were not required for admission into the occupational therapy program at Wayne State University. Therefore, results of the correlation analysis for GRE and the station score were based on the nine individuals, or approximately 24.3% of the sample population. Table 14 displays descriptive statistics for the GRE scores and the station score.

Table 14.

Descriptive Statistics for GRE and Station Scores

			Bootstrap[a]			
					95% Confidence Interval	
Variable		Statistic	Bias	Std. Error	Lower	Upper
Verbal GRE	Mean	145.33	-.02	1.68	141.89	148.44
	Std. Deviation	5.268	-.403	.966	2.819	6.540
	N	9	0	0	9	9
Quantitative GRE	Mean	144.11	-.02	1.73	140.33	147.11
	Std. Deviation	5.442	-.572	1.609	1.936	7.517
	N	9	0	0	9	9
Analytic GRE	Mean	3.722	.001	.252	3.222	4.167
	Std. Deviation	.7949	-.0820	.2261	.2500	1.0833
	N	9	0	0	9	9
Total Traditional	Mean	30.33	-.03	3.10	23.78	35.89
	Std. Deviation	9.747	-.828	2.322	3.005	12.500
	N	9	0	0	9	9
Total Open-ended	Mean	150.500	.056	7.547	133.947	163.332
	Std. Deviation	23.9726	-2.6078	7.3506	6.8702	33.1227
	N	9	0	0	9	9
Total Overall Score	Mean	180.8333	.0257	7.3518	165.8347	194.2208
	Std. Deviation	23.31443	-2.10475	5.71240	10.67955	31.48807
	N	9	0	0	9	9

a. Unless otherwise noted, bootstrap results are based on 10000 bootstrap samples

Results of the Pearson correlation analysis can be seen in Table 15. None of the GRE scores was correlated to any other measure examined in this study (p = .059 - .999). The only statistically significant finding was the total score for the open-ended stations was significantly correlated with the total overall score on all stations (p = .001).

Table 15.

Correlations between Three Types of GRE and Total Station Scores

Variable		Verbal GRE	Quantitative GRE	Analytic GRE	Total Traditional	Total Open-ended	Total Overall Score
Verbal GRE	Pearson Correlation	1	.648	-.020	.662	.000	.276
	Sig. (2-tailed)		.059	.959	.052	.999	.472
	N	9	9	9	9	9	9
Quantitative GRE	Pearson Correlation	.648	1	-.411	.155	.008	.073
	Sig. (2-tailed)	.059		.272	.691	.983	.852
	N	9	9	9	9	9	9
Analytic GRE	Pearson Correlation	-.020	-.411	1	-.204	.298	.221
	Sig. (2-tailed)	.959	.272		.598	.435	.567
	N	9	9	9	9	9	9
Total Traditional	Pearson Correlation	.662	.155	-.204	1	-.270	.141
	Sig. (2-tailed)	.052	.691	.598		.482	.718
	N	9	9	9	9	9	9
Total Open-ended	Pearson Correlation	.000	.008	.298	-.270	1	.915**
	Sig. (2-tailed)	.999	.983	.435	.482		**.001**
	N	9	9	9	9	9	9
Total Overall Score	Pearson Correlation	.276	.073	.221	.141	.915**	1
	Sig. (2-tailed)	.472	.852	.567	.718	**.001**	
	N	9	9	9	9	9	9

** Correlation is significant with a Bonferroni adjustment (p = .05/5 = .01).

Item Difficulty and Item Discrimination

Because the reliability between the open-ended and traditional stations was poor, based upon Cronbach's alpha results, the decision was made to analyze item difficulty and item discrimination for the two types of stations separately. For the traditional stations, the maximum score was 9, thus a mean (P) of 4.5 was considered a 'good' rating scale. Results are compiled in Table16. Three of the traditional stations had good item

difficulty; positive self-concept (*P* = 4.81), navigate systems / culture (*P* = 5.08) and community / interpersonal (*P* = 4.43). Community service was too difficult (*P* = 2.68), and realistic self-appraisal (*P* = 7.00) and leadership (*P* = 7.14) were too easy.

The only item from the traditional stations which appeared to discriminate between test takers was realistic self-appraisal (*Rpbis* = 0.26). Positive self-concept (*Rbpis* = 0.04), communication / interpersonal (*Rpbis* = 0.02), and navigate systems / culture (*Rpbis* = 0) had low or no discrimination. Leadership (*Rpbis* = -0.05) and community service (*Rpbis* = -0.01) may be considered poor items due to their negative Pearson point-serial correlation values.

Table 16.

Item Difficulty and Discrimination for Traditional Stations

Statistic	Mean	SD	Min Score	Max Score	*P*	Total *Rpbis*
Scored Items	31.14	6.20	15	43	5.19	0.04
Positive self-concept	4.81	2.54	0	9	4.81	0.04
Realistic self-appraisal	7	1.94	0	9	7.00	0.26
Navigate systems / culture	5.08	2.95	0	9	5.08	0
Leadership	7.14	2.31	2	9	7.14	-0.05
Community service	2.68	2.81	0	9	2.68	-0.01
Communication / interpersonal	4.43	1.94	2	9	4.43	0.02

Notes: P = item mean (difficulty), Total Rbpis = item point-biserial correlation (discrimination)

For the open ended stations, the maximum score was 5 for each item, making a mean (*P*) of 2.5 as a ‘good’ rating scale for item difficulty. The stations of communication / interpersonal (P = 2.54), community service (*P* = 2.74), and leadership (*P* = 3.14) were considered to have appropriate difficulty (see Table 17). The remaining three items might

be considered easy (*P* = 3.51 – 3.70). All six items had appropriate discrimination (*Rbpis* = 0.15 – 0.56).

Table 17.

Item Difficulty and Discrimination for Open-ended Stations

Statistic	Mean	SD	Min Score	Max Score	P	Total Rpbis
Total Score	19.30	2.30	11	22	3.22	0.44
Positive self-concept	3.70	0.46	3	4	3.70	0.44
Realistic self-appraisal	3.51	0.65	1	4	3.51	0.43
Navigate systems / culture	3.68	0.75	2	5	3.68	0.56
Leadership	3.14	0.75	2	4	3.14	0.38
Community service	2.73	0.51	1	3	2.73	0.66
Communication / interpersonal	2.54	0.51	2	3	2.54	0.15

Note: P = item mean (difficulty), Total Rbpis = item point-biserial correlation (discrimination)

Bias

Tables 18, 19, and 20 show results of the Fisher exact tests for the traditional stations, open-ended stations, and total overall score. There were no statistically significant differences between scores on the three outcomes measures based upon sex (*p* = .394 - .925), ethnicity (*p* = .029 – 1.00), Pell grant status (*p* = .394 - .694), family college history (*p* = .124 - .948), or income level (*p* = .070 - .477).

Table 18.

Fisher's Exact Test Results for Final Score on Traditional Stations

Group	Category	n	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Fisher's exact test value	Exact Sig.* (2-sided)
Sex	Male	5	1	3	1	0	2.466	.476
	Female	32	8	10	5	9		
Race	Minority	10	2	4	2	2	.659	1.000

	Caucasian	27	7	9	4	7		
Pell Grant	Yes	9	2	4	0	3	2.817	.450
	No	27	7	9	6	5		
College	Not 1st generation	29	6	12	3	8	5.232	.124
	1st generation	8	3	1	3	1		
Income	< $47,500	7	4	1	1	1	11.626	.170
	$47,500 - $70,000	8	1	4	1	2		
	$70,001 – $100,000	5	2	1	2	0		
	>$100,000	10	1	6	0	3		

*Significance with Bonferroni adjustment set at (p = .05/5 = .01)

Table 19.

Fisher's Exact Test Results for Final Score on Open-ended Stations

Group	Category	n	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Fisher's exact test value	Exact Sig.* (2-sided)
Sex	Male	5	2	1	0	2	3.009	.394
	Female	32	7	10	9	6		
Race	Minority	10	3	6	0	1	8.141	.029
	Caucasian	27	6	5	9	7		
Pell Grant	Yes	9	2	4	3	0	3.009	.394
	No	27	7	6	6	8		
College	Not 1st generation	29	7	8	7	7	.784	.948
	1st generation	8	2	3	2	1		
Income	< $47,500	7	0	3	2	2	13.881	.070
	$47,500 - $70,000	8	2	2	4	0		
	$70,001 – $100,000	5	0	2	1	2		
	>$100,000	10	5	0	2	3		

*Significance with Bonferroni adjustment set at (p = .05/5 = .01)

Table 20.

Fisher's Exact Test Results for Total Overall Score

Group	Category	n	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Fisher's exact test value	Exact Sig.* (2-sided)
Sex	Male	5	2	1	1	1	1.052	.925
	Female	32	7	9	8	8		
Race	Minority	10	3	3	4	0	5.301	.163
	Caucasian	27	6	5	9	7		
Pell Grant	Yes	9	2	3	3	1	1.849	.695
	No	27	7	7	5	8		
College	Not 1st generation	29	7	7	7	8	1.163	.946
	1st generation	8	2	3	2	1		
Income	< $47,500	7	0	2	3	2	8.944	.447
	$47,500 - $70,000	8	2	3	3	0		
	$70,001 – $100,000	5	1	2	0	2		
	>$100,000	10	3	3	1	3		

*Significance with Bonferroni adjustment set at (p = .05/5 = .01)

CHAPTER 5 - DISCUSSION

The aim of this study was to examine the psychometric properties of the *Computer-based Assessment of Non-cognitive Attributes of Health Professionals* (CANA-HP). Three research questions were delineated which focused on different aspects of the CANA-HP related to internal consistency (reliability), inter-rater reliability, construct validity, item difficulty and discrimination, and bias of the instrument toward individuals from a variety of backgrounds.

In measurement methodology, assessment of the same attribute should demonstrate homogenous results or internal consistency (Portney & Watkins, 2020). Reliability analysis indicated the stations of the CANA-HP had low correlation ($\alpha = 0.492$). One possible explanation is the two types of stations (traditional and open-ended) might be measuring different traits. Based on this rationale, the two types of stations were analyzed separately. For the traditional stations (multiple choice questions), the reliability worsened when the six questions were examined against each other ($\alpha = 0.091$), indicating no internal consistency. However, the open-ended stations were found to have moderate correlation ($\alpha = 0.706$) when examined independently from the traditional stations. Note these magnitudes cannot be interpreted on the usual scale for practical purposes from zero to 1. (Theoretically, they can be negative). The statistical engine of internal consistency reliability is the Pearson correlation, which will attenuate (shrink) as the number of items decreases.

The reliability obtained for the six open-ended stations are similar to results found by Dore et al. (2009), developers of CASPeR. It included eight case vignettes and four self-descriptive questions designed to assess non-cognitive attributes of students

applying to medical school. The open-ended stations of the CANA-HP included six professional dilemmas which could be encountered by students applying for admission into one of four health science programs. The overall reliability for the typed CASPeR was 0.72 (Dore et al., 2009) compared with 0.71 for the CANA-HP. (CASPeR had both a typed and audio version.) Items with strong internal consistency should only show moderate correlations (between 0.70 and 0.90 when there are a large number of items). When correlation gets too high, there is a concern the items being measured may be redundant with a potential for limitations in the content validity (Portney & Watkins, 2020), which could be ameliorated by reduction with a factor analysis. Hence, the open-ended stations of the CANA-HP had a minimally acceptable level of internal consistency, without being redundant.

The six traditional stations of the CANA-HP contained multiple choice questions in which the candidate chose the three most correct answers from a pre-determined list of seven multiple choice options. Scoring was completed by survey software after candidates submitted their responses. Patterson et al. (2016) used a similar type of formatting, described as situational judgement tests, for students applying to medical school. Candidates in this latter study selected responses from pre-determined options which included multiple choice, ranking, or single best answer. Patterson et al. reported an internal consistency which ranged from $\alpha = 0.43 – 0.94$, compared with the internal consistency of $\alpha = 0.091$ for the traditional stations in this study. The advantages of the traditional, multiple-choice format is the design is cost-effective and efficient for programs to screen multiple candidates. In comparison, the open-ended stations obtained in this study took an average of 6 minutes and 29 seconds to grade. For programs with a large

number of applicants, faculty time may be spent grading applicants who never enter the program. In 2017, for example, the physician assistant program at Wayne State University had 350 applicants. If only one reviewer scored each applicant, over 37 hours of faculty time would be dedicated to this process. A program would need to weigh the decision to outsource grading of open-ended stations to outside agencies, like the developers of CASPeR, or maintain internal control and cost reduction by developing or refining a tool such as the CANA-HP. For programs such as occupational therapy, no vendor currently offers a tool like CASPeR. Therefore, an internal methodology would have to be developed.

There are several suggestions which may improve the low internal consistency found with the traditional stations. A Spearman Brown prophecy formula was computed, and an additional 144 stations would be needed to achieve a minimal Cronbach's alpha level of 0.70. Because the traditional stations are computer scored, additional questions would not add to faculty workload. The additional time expenditure would be for the initial development of the extra questions. Therefore, one suggestion is to increase the number of stations.

A second consideration is to run the sample with a larger number of students and with applicants from other health care professions (nurse anesthesia, physical therapy, and physician assistants). Applicants to the occupational therapy program may not be representative of other health care professionals. Applicants in Patterson et al. (2016), for example, were applying for medical school and the differences between the two student populations may have resulted in differences in the reliability between the two methodologies. Consideration also might be given to changing the format of the traditional

stations used in this study. Participants were asked to choose the best three answers from a list of seven choices. Perhaps a single best answer format may improve internal consistency.

Interrater reliability was assessed for the CANA-HP and in general was found to range from moderate (*r* = 0.67) to excellent (0.91) for the two raters (Koo & Li, 2016). CASPeR was reported to have a general interrater reliability of r = 0.81, which is consistent with these findings. However, scoring of the CANA-HP took only 6 minutes and 29 seconds for all six stations (roughly one minute, five seconds per station), compared with 24 minutes to score the 12 stations of the CASPeR tool. The CANA-HP is less time intensive based on these results.

According to Koo & Li (2016), the 95% confidence interval should be reported with ICC values. The CANA-HP reliability for each station was 0.364 – 0.953 which could be interpreted as poor to excellent reliability for the raters. The two constructs obtained in the current study with the lowest interrater reliability were positive self-concept and realistic self-appraisal. Positive self-concept measures an applicant's ability to express confidence and strength of character. The raters only differed by an average of one point during the training period, the lowest of all the differences observed between the raters. However, there was fluctuation in ratings during this period, with both the experienced and novice rater alternating as to who provided the higher or lower score for an applicant.

Realistic self-appraisal relates to an applicant's ability to self-develop, apply critical thinking, and recognize a need to broaden his/her individuality. It had the most inconsistency between the two raters during the training period. The raters had a 6.33 mean point difference when looking at these items during training. The rater with more

academic experience consistently rated candidates lower for eight of the nine reviews.

Several considerations for improving interrater reliability are recommended for future studies. It may be beneficial to find raters who are homogenous in nature, such as faculty who review students applying to a health care profession. In this study, one rater was a faculty member in a health profession for 14 years. The other rater was enrolled in college studying secondary education, but was not a health professional nor an experienced teacher. Higher reliability was reported between raters who were homogenous in background and experience (Follman & Anderson, 1967).

Another recommendation is to consider an order effect which may be present in grading. Because both positive self-concept and realistic self-appraisal were the first items scored, each rater may have become more consistent with scoring over time. Randomly changing the order in which items are reviewed for each candidate may negate the potential impact of order. In addition, rating all candidates on one construct at a time, rather than rating the complete rubric for one candidate, may be beneficial. Consideration should be given to providing additional clarification and / or descriptions to the rubric used to score these two stations (Appendix C).

Because the CANA-HP was designed to measure non-cognitive attributes of applicants to health care professions, the hypothesis was this novel methodology would not be significantly correlated to traditional measures of cognitive abilities such as GPA and GRE scores. All three of the GPA scores (science, non-science, and undergraduate) were significantly correlated to each other. Because GPA measures componential and analytical thinking (Kalsbeek, 2013), and GPA scores scaffold on top of each other, this is an expected finding. The final score on the traditional stations was not significantly

correlated to GPA scores, supporting the hypothesis. However, the final score on the open-ended stations and total overall scores were significantly correlated to non-science GPA (p = .002 and p = .008 respectively). This represents a medium effect size (r = 0.496 open-ended stations, r = 0.429 total overall score). This finding does not support the hypothesis, because the CANA-HP was designed to measure non-cognitive attributes and GPA is considered to reflect cognitive thinking.

Science GPA is composed of classes like biology, math, chemistry, and physics. For many health care programs, science GPA is a pre-requisite for admission into the program. Science classes have been reported to be analytical, residing in the cognitive realm (Kalsbeek, 2013). Analytical thinking involves interpreting information in well-defined and unchanging contexts, and students are often tested using standardized tests (such as multiple choice). However, non-science GPA is comprised of all other classes a student takes. Classes in the non-sciences can be varied and broad, and could include courses in exercise, dance, foreign language, music, and liberal arts. They contain a mixture of both analytical and experiential / creative learning, a subset of intelligence where individuals learn to interpret novel tasks. Kalbeek (2013) reported individuals from non-traditional backgrounds may use this latter type of intelligence during initial exposure to new subject matter.

If candidates applying to health science programs took a number of classes which encouraged experiential / creative learning, non-science GPA may in part measure a portion of non-cognitive abilities. Thus, non-science GPA may represent another outcome which has non-cognitive dimensions associated with it. The correlation between non-science GPA and the open-ended stations may be valid, as both could be identifying non-

cognitive attributes of applicants.

None of the health care program formulas at Wayne State University consider non-science GPA as part of the scoring rubric for admission. However, this variable should be further explored as a potential measure of non-cognitive attributes. CASPeR and CMSENS, tools designed to assess non-cognitive attributes of applicants to medical school, have been analyzed for correlation to cognitive outcome measures commonly used for medical school admission (such as the Medical College Admission Test (MCAT)) and medical licensure examinations (Part I and Part II). The relationship between these two tools and non-science GPA scores has not been reported. Therefore, the correlation between non-science GPA and the CANA-HP stations may be novel and should be investigated further.

The greatest portion of the total overall score was compromised of the scores on the open-ended stations (mean open-ended station scores = 154.4 versus mean traditional station scores = 30.7). Therefore, it is logical if the open-ended stations were significantly correlated to non-science GPA, the total overall score would have similar results.

GRE scores were not correlated to each other or any of the total scores on the CANA-HP stations. This supports the hypothesis of this study. The only statistically significant finding was the total score for the open-ended stations was significantly correlated to the total overall score on all stations. As previously mentioned, this might be attributed to the overall percent contribution the open-ended stations had on the total overall station scores.

The lack of correlation between CANA-HP and GRE scores is in contrast to results

reported by Dore et al. (2009) in which CMSENS was found to have a significant correlation to the MCAT with a small effect size (r = .018 – 0.28). The MCAT is a computer based assessment which tests knowledge of physical and biological sciences, verbal reasoning and writing skills, similar to the GRE with its verbal, quantitative and analytical components. The MCAT is taken prior to admission to medical school, much like the GRE is taken prior to admission to graduate school. Therefore, both tests might be thought to examine cognitive skills of test takers. The reasons for the contrasting results between the CANA-HP and CMSENS are not known. There are some basic differences between the two tools which might have impacted the results. The CANA-HP has six stations compared to 12 stations of the CMSENS. Raters of the CMSENS used a nine-point Likert scale to score respondents. Raters for the CANA-HP used a 5 point Likert Scale. Finally, the CMSENS contained 12 vignettes, four which were self-descriptive in nature (such as what do you do best?) and eight which showed a video of a generic ethical scenario (such as how to respond to an ethical dilemma when working as a cashier). The CANA-HP, on the other hand, was designed using only six stations. All six stations were specific to the medial field (not profession specific) and posed questions to tease out the six constructs the tool was designed to measure (see Appendix B).

Bivariate correlations were conducted between CASPeR and outcomes measures on the medical licensure examinations for both Part I and Part II (Dore et al, 2017). In general, CASPeR was correlated with the professional domains on this exam with a medium to large effect size (r = 0.30 – 0.50). There was no correlation between the cognitive portion of the test and CASPeR outcomes. To further validate the CANA-HP, predictive validity needs to be conducted between CANA-HP results and non-cognitive

outcomes for students accepted into a health care program. Consideration should be given to outcomes which are not specifically related to cognitive measures such as practical examinations, objective structured clinical examinations, and clinical education performance. This is an area of need for the current methodology.

Item difficulty, item discrimination, and bias of the overall methodology was also considered. The open-ended stations had a minimally acceptable level of reliability, and appropriate discrimination (*Rpbis* = .015 – 0.56). Three of the stations had appropriate difficulty (*P* = 2.54 – 3.14, target *P* = 2.5), but the remaining three stations might be considered too easy. Continued analysis on these stations is recommended by increasing the sample size and comparing results between applicants of different health care professions.

The traditional multiple choice stations need additional refinement. The reliability for these stations was low. In addition, only one station was found to discriminate between test takers, realistic self-appraisal (*Rbpis* = 0.26). Three of the remaining stations had no or low discrimination and two might be considered ‘poor’ items. Half of the stations had good item difficulty (*P* = 4.43 – 5.08, target *P* = 4.50), two stations would be considered easy, and one was too difficult. As mentioned previously, several changes to these items are recommended. Increasing the number of stations might allow for increased internal consistency, analysis of reliability by domain, and may change item bias and discrimination results. Changing the format from selecting the three best options to choosing the single best answer may also impact these findings. Increasing the sample size to include applicants to other health professions may positively impact future studies. Item discrimination and item difficulty cannot be compared to either the CMSENS or

CASPeR tools because these findings were not reported.

There was no statistically significant difference on the three outcome measures (total score traditional stations, total score open-ended stations, and total overall score) for the variables of sex (male, female), ethnicity (minority, Caucasian), Pell grant status (recipient, not recipient), family college history (1st generation in college, not 1st generation in college), and income level (broken into quartiles). Hence, the CANA-HP is not biased toward any of the variables mentioned. These results should be interpreted with caution, because the sample sizes for each variable was small and had to be collapsed for analysis. A larger sample is needed to increase the number of individuals in each variable category. Because the purpose of this study is to find a methodology to measure non-cognitive abilities of applicants, it is imperative the tool is not biased toward any group from a non-traditional background. For example, Rees et al. (2016) reported the Multiple Mini-Interview, commonly used in medical school admissions, disadvantaged rural applicants and urban bias should be explored by programs which use the tool.

Limitations of the Study

The participants in this study were a small sample of convenience, limited to applicants to one health profession program at the university where the study was conducted. Therefore, these results may not be generalizable to applicants at other universities or within other health care professions. In particular, the number of applicants to the occupational therapy at the time this study was conducted (*N* = 38) was approximately one half of the applicants who normally applied to the program (*N* = 74). One suspected reasons was the program changed the terminal degree from a Bachelor's to a Master's Degree at the time of this study. This degree change would increase overall

tuition costs for students (undergraduate versus graduate tuition), and would require more stringent criteria to stay in the program (no grade was accepted below a C). Similar decreases in applicants were seen in other health professions at this university when a program changed the terminal degree. Applicants who still choose to apply to the occupational therapy program, despite the change in degree, may not represent the applicants seen in previous years or by other health care programs.

Although open-ended situational judgment tests were reported to have adequate internal consistency, little research has been conducted using pre-selected multiple-choice options (traditional stations). Patterson et al. (2016) reported internal consistency which ranged from $\alpha = 0.43 – 0.94$ for multiple choice questions with pre-determined options (situational judgment tests). However, these items were difficult to design and significant expertise was required to build a reliable and valid situational judgment test. Although the consultants and primary researcher in this study had a long history of experience in their health professions (minimum of 20 years of experience in the profession), they had no previous knowledge writing questions in the format of a situational judgment test. This lack of experience may have impacted the results leading to the low item discrimination and low reliability for the traditional stations of the CANA-HP.

The constructs used in this study may have been difficult to measure as none had a gold standard methodology to use as a criterion reference. In addition, each construct had only one question for both the traditional and open-ended stations. Domain reliability could not be obtained, and as shown by a Spearman Brown prophecy formula, more questions could have resulted in higher reliability scores. The open-ended questions had

a higher weighted value which might have influenced the total overall scores for the CANA-HP. Future studies should add more questions and increase the weight given to the traditional stations.

Future studies could examine CANA-HP scores and outcomes related to non-cognitive attributes desired in health care providers. Some examples of outcomes for predictive validity might include practical examinations, objective structured clinical examinations, and clinical experiences. It is important to analyze how non-cognitive attributes impact a student's ability to provide patient care. The raters in this study were not homogenous and the lack of homogeneity may have impacted findings. Reviewers who are similar to the individuals who will ultimately be reviewing applicants into the program may improve the internal consistency reliability.

Implications for Future Research and Practice

Assessment of the non-cognitive attributes of applicants to health care programs has become increasingly sought after by many health care programs as professions look for ways to increase diversity among working clinicians. Although the open-ended scenarios of the CANA-HP were found to have minimally acceptable reliability, adequate item discrimination, and adequate item difficulty, further work is needed to refine these stations. Consideration should be given to increasing the number of questions for each construct to further enhance internal consistency, as well as increasing the sample size used in analysis. For the traditional stations much work still needs to be done. More questions need to be developed, particularly with the help of experts who have prior experience writing these questions. Faculty members who attempt to develop their own situational judgment tests should seek experts in the field to assist with initial question

development and then run psychometric analyses on the developed methodology prior to using the new tool in the actual application process. Analyses of different formats of multiple choice questions should be considered as this study only examined selecting the three best responses. A single response or ranking may provide better discrimination and internal consistency.

In considering adding either traditional or open-ended stations to the application process there should be heavy weight given the time factor associated with each type of methodology. When a multiple choice situational judgment test is well designed, it can be cost effective and easy to administer to a large number of candidates despite the initial time investment (Patterson et al., 2016). In general, the open-ended scenarios used in this study will be time intensive for faculty to review. If the decision is made to pay to have the applicant reviewed, the cost may be prohibitive to the program and/or applicant. Many of the traditional formats used by programs to test non-cognitive attributes of candidates, such as the structured interview or personal statement, are labor intensive to faculty. This may be one of the biggest complaints when considering switching to a process which examines the non-cognitive attributes of program applicants. When pushed to offer a more holistic method for program admissions, program administrators may decide to use time intensive methodologies due to familiarity with these tools. In addition, programs may not have the time or faculty expertise to try newer methodologies such as situational judgment tests.

Non-science GPA may be another outcome for further exploration by programs. It may represent a potential measure of non-cognitive attributes for students applying for admission into health care programs. The traditional stations were significantly correlated

to non-science GPA. Non-science GPA has not been studied as a potential measure of non-cognitive attributes among health care applicants. At the university where this study was conducted, non-science GPA is not included in the scoring rubric for any of the health professions. Science GPA, however, is used in scoring as it has been felt to be a better predictor of ability to successfully complete the program and pass licensing board examinations. However, science GPA measures cognitive attributes.

Conclusion

The CANA-HP remains a work in progress. Initial results support the hypothesis of no correlation with standardized cognitive assessments (GRE and GPA scores). The one exception was non-science GPA which was significantly correlated to the total open-ended scores and total overall score, and should be further examined. The six open-ended scenarios had minimally adequate internal reliability, and adequate item discrimination / difficulty. The traditional multiple choice questions need further refinement as these six scenarios had low reliability and discrimination. Homogenous raters may improve interrater reliability. Predictive validity of this methodology is needed.

APPENDIX A: IRB EXPEDITED APPROVAL

WAYNE STATE UNIVERSITY

IRB Administration Office
87 East Canfield, Second Floor
Detroit, MI 48201
Phone: (313) 577-1628
www.irb.wayne.edu

CONCURRENCE OF EXEMPTION
IRB-19-12-1558-B3 Expedited/Exempt Review-EXEMPT

DATE:	January 14, 2020
TO:	Maher, Sara, Education Dean
	Hill, William, Administration & Organization Stud
FROM:	Millis, Scott, Professor, B3 Expedited/Exempt Review
PROTOCOL TITLE:	Pilot Evaluation of the Computer-Based Assessment of Non-Cognitive Attributes of Health Professionals (CANA-HP).
FUNDING SOURCE:	NONE
PROTOCOL NUMBER:	IRB-19-12-1558

The above-referenced protocol has been reviewed and found to qualify for Exemption according to category 2

The following attachments and consent/assent documents have been reviewed and approved by the IRB.

Notes:
Note to PI: This application has been given a Status Check-In Date. Please submit a Status Update Report for this project by 01/13/2022. The Minimal Risk Status Update Form is available on the IRB's website. Modifications/changes to the research project will need to be submitted via an amendment to the WSU IRB.

Protocol/Proposal/Dissertation (dated 01/08/2020)

Research Information Sheet (dated 12/2019)

Initial Email from Admission Committee Chair

Data Collection Tool (1): (I) CANA-HP Survey

Attachments

APPROVED Research Information Sheet 1.10.20
APPROVED Maher (Initial email) (1)
APPROVED CANA-HP Rev #7 (includes demographics) (1)
CANA-HP Rev #7 (includes demographics)
Maher (Initial email)
Proposal for IRB (Revised 1.10.20)

* Exempt protocols do not require annual review by the IRB, however you may have been granted a Status Check-In Date.

* All changes or amendments to the above-referenced protocol require review and approval by the IRB BEFORE implementation.

* Adverse Reactions/Unanticipated Problems AR/UP must be submitted on the appropriate form within the time frame specified in the IRB. In the event of an unexpected problem use the Unanticipated Problem Report Form.

Note: Studies conducted at DMC sites or DMC medical record used for affiliate review Authorized DMC personnel have been added to this submission under Personnel Information "Other".
Administration Office Policy www.irb.wayne.edu/policies-human-research

APPENDIX B: CANA-HP STATIONS

TRADITIONAL STATIONS

Question #1: Positive Self -Concept
While caring for a patient as a student in a health care program, you made a treatment error which you did not recognize at the time. The error resulted in no harm to the patient, and there was no one in the area who saw your mistake. The patient did consent to care, and is not aware there was any problem. Several days have passed and you will not see the patient again.

Choose the THREE most appropriate responses.

- ○ A. Inform your preceptor / clinical instructor of the error and ask for advice on how to proceed.
- ○ B. Document what occurred in the patient chart and include the patient's response to the error.
- ○ C. Continue today's schedule as planned and make no reference to the error.
- ○ D. Complete a clinical incident form and notify risk management of the error.
- ○ E. Find a colleague and discuss specific details to determine best actions moving forward.
- ○ F. Inform the patient of the error and discuss potential side effects.
- ○ G. Call the recipient rights advisor and ask for advice on how to proceed.

ANSWER: Correct: ABF (Incorrect: GEDC (rank ordered))

This question deals with the test taker assuming responsibility for his / her actions while demonstrating strength of character consistent with a positive self-concept.

A. *Inform your preceptor / clinical instructor of the error and ask for advice on how to proceed* is a correct option. In this scenario, the test taker is electing to admit the error by notifying the immediate supervisor. In addition, the test taker is able to recognize there are many different responses to a treatment error based upon

the health care system one works under. A student would not be expected to have full system knowledge and should ask for help. (+3 points)

B. *Document what occurred in the patient chart and include the patient's response to the error* is another correct option. By documenting the error, the test taker is assuming responsibility for actions. Furthermore, through documentation the health provider is identifying how the patient responded to the treatment should it be questioned later. (+3 points)

F. *Inform the patient of the error and discuss potential side effects* is another correct option. The test taker in this scenario is again taking responsibility for actions. In addition, the test taker has alerted the patient to potential for harm. The risk in this scenario is the health care system may want to be aware of such situations before patients are informed. (+3 points)

G. *Call the recipient rights advisor and ask for advice on how to proceed* is a neutral option. While the test taker has identified an error and is seeking help, the recipient rights advisor handles issues where a patient's rights have been violated. There is no clear indication in the stem the treatment error resulted in a violation of patient rights because the patient did consent to treatment. (0 points)

E. *Find a colleague and discuss specific details to determine best actions moving forward* is an incorrect option. While the test taker is attempting to learn what the best action is in the situation, sharing information about a patient with a colleague who may or may not be involved in care of the patient is a violation of patient privacy. (-1 points)

D. *Complete a clinical incident form and notify risk management of the* error is an incorrect option. In this scenario, the test taker is assuming an incident occurred. However, a **clinical incident** is any unplanned event which causes, or has the potential to cause, harm to a patient. The case presented does not meet this criteria and has the potential to waste time and money into an investigation. (-2 points)

C. *Continue today's schedule as planned and make no reference to the error* is an incorrect options. In this scenario, the test taker is attempting to cover up the action which occurred. This behavior does not demonstrate trying to understand or navigate a system, but rather to protect self from potential harm from an incorrect treatment. (-3 points)

Question #2: Realistic Self-Appraisal

You have been asked to work with a patient with whom you previously had difficulty providing care. The patient instantly recognizes you and states "I don't want you anywhere near me". "You don't know what you are doing and make me uncomfortable".

Choose the THREE most appropriate responses.

○ A. Reassure the patient you are competent in your patient care skills and can work with them.

○ B. Apologize to the patient for previous care and discuss the plan for today.

○ C. Inform the patient the next available appointment with another practitioner is two weeks away.

○ D. Explain to the patient no one else is available to provide care at this time so care must be provided by you.

○ E. Discuss with the patient the plan of care to discover what makes the patient uncomfortable.

○ F. Exchange patients with a colleague who works in the same treatment area.

○ G. Listen to the patient and then make minor revisions to today's plan of care.

ANSWER: Correct: ABE (Incorrect: GFCD (rank ordered))

This question deals with the test taker recognizing and accepting personal strengths and deficits. The test taker's response should demonstrate self-development, ability to apply critical thinking, and ability to broaden treatment scope.

A. *Reassure the patient you are competent in your patient care skills and can work with them* is a correct option. The option addresses the issue of competence, works to make the patient comfortable with the health care provider, and directly addresses the issue at hand. (+ 3 points)
B. *Apologize to the patient for previous care and discuss the plan for today* is another correct option. Apologizing for prior treatment shows the patient the practitioner accepts responsibilities for actions. The patient may be more likely to allow current care. However, this response is often best accompanied by reassurance of the current abilities of the health care provider. (+ 3 points)
E. *Discuss with the patient the plan of care to discover what makes the patient uncomfortable* is another correct option. The test taker in this option is exhibiting

a willingness to know what makes the patient uncomfortable and is willing to learn from the patient. (+3 points)

G. *Listen to the patient and then make minor revisions to today's plan of care* is a neutral option. While listening skills show empathy, the test taker is still proceeding with the plan of care, making only minor revisions. This response does not acknowledge the patient's distress nor does it acknowledge responsibility for actions. (0 points)
F. *Exchange patients with a colleague who works in the same treatment area* is an incorrect option. While it does address the patient discomfort, the practitioner is not accepting responsibility for actions. Furthermore, with this response the test taker is avoiding the opportunity to self-reflect and learn more about what has made the patient uncomfortable. (-1 point)
C. *Inform the patient the next available appointment with another practitioner is two weeks away* is an incorrect option. This is an example of coercing a patient to consent to being treated by the practitioner. It forces the patient to delay care and does not directly address the situation at hand. (-2 points)
D. *Explain to the patient no one else is available to provide care at this time so care must be provided by you* is another incorrect options. Not only is this an example of coercing a patient to consent to treatment, but the patient is not given any choices for his/her own plan of care. (-3 points)

Question #3: Able to navigate systems

You have been accepted into a health profession program. You are currently a student on a hospital rotation, completing an initial evaluation for a patient you are scheduled to care for tomorrow. During the history and physical, a technician from x-ray comes into the room and states the patient needs to be taken to the diagnostic center for an immediate x-ray. The technician begins gathering the patient's belongings and proceeds to wheel the patient out of the room.

Choose the THREE most appropriate responses.

○ A. Ask your preceptor / clinical instructor for advice on how to handle the situation.

○ B. Arrange to speak to the technician later to discuss your working relationship.

○ C. Walk with the patient and continue to gather the remaining items for you history.

○ D. Inform the technician you will be done shortly and please wait in the waiting area.

○ E. Call the technician's supervisor to reschedule the x-ray for a later time period.

○ F. Instruct the nurse to immediately call the physician for clarification.

○ G. Document the information which has been gathered and finish the evaluation later.

ANSWER: Correct GAB (Incorrect: DEFC (rank ordered))

This question deals with the test taker exhibiting a realistic view of working in a health system. The test taker is committed to improving the system and yet, is not hostile to working within it.

G. *Document the information which has been gathered and finish the evaluation later* is a correct response. The test taker recognizes a hospital system involves a lot of moving pieces and working around scheduled (or unscheduled) tests is part of the system. The test taker should recognize the importance of documenting what has already occurred, and the evaluation can resume at a later time. (+ 3 points)

A. *Ask your preceptor / clinical instructor for advice on how to handle the situation* is a correct response. Here the test taker recognizes diagnostic tests are difficult to reschedule. However, the test taker also is not sure of how to deal with these situations in the future, so discussing with the preceptor / clinical instructor will help to better navigate the system in the future. (+3 points)

B. *Arrange to speak to the technician later to discuss your working* relationship is a correct option. Here the test taker recognizes diagnostic tests are difficult to reschedule and accommodates the test. However, the test taker also is not sure of how to deal with these situations in the future, so discussing with the technician will help to better navigate the system in the future. (+3 points)

D. *Inform the technician you will be done shortly and please wait in the waiting* area

is a neutral response. While the test taker is exhibiting a lack of knowledge about hospital systems, the test taker has not violated confidentiality and has demonstrated lack of knowledge to the technician only. The technician will most likely explain immediately to the test taker why the patient must be taken for imaging. (0 points)

E. *Call the technician's supervisor to reschedule the x-ray for a later time* period is an incorrect option. Not only is the test taker demonstrating a lack of understanding of hospital systems, he/she has involved management and gone above the head of a colleague within the system before speaking to the colleague. (-1 point)

F. *Instruct the nurse to immediately call the physician for clarification* is an incorrect option. Not only is the test taker demonstrating a lack of understanding of hospital systems, he/she has involved two additional individuals in this situation, the nurse and physician. This behavior demonstrates a lack of knowledge for who to contact within the system. (-2 points)

C. *Walk with the patient and continue to gather the remaining items for you history* is an incorrect response. This is a direct violation of patient rights to confidentiality of treatment. The test taker is displaying complete lack of knowledge of systems or patient rights. (-3 points)

Question #4: Leadership

A severe ice storm has caused a major accident on several freeways resulting in numerous injuries and several deaths. The storm has affected power to the hospital causing the hospital to rely on back-up generators for essential functions. Your day shift is scheduled to end in 30 minutes and you are responsible to get to the elementary school to pick up your child. Your parents and spouse are not in town. The area supervisor has informed you of the requirement to stay at the hospital until power is restored.

Choose the THREE most appropriate responses.

○ A. Ask a colleague from the local area, who is not employed by the hospital, to come to the hospital to cover for you.

○ B. Inform your supervisor of your responsibilities for your child and leave the hospital.

○ C. Call the school and have the child placed in after school care until you can get there.

○ D. Arrange to have your child cared for by a trusted neighbor until you can leave work.

○ E. Notify your supervisor of your child's situation and ask to leave as soon as possible.

○ F. Stay at the hospital until such time as it is absolutely necessary to get your child.

○ G. Ask permission to speak to the supervisor's boss to discuss the need to leave the hospital.

ANSWER: Correct DCE (Incorrect: GFAB (rank ordered))

This question deals with the test taker demonstrating leadership in any area of background. By his/her actions, the test taker should show leadership responsibility for both patient care and his/her children.

D. *Arrange to have your child cared for by a trusted neighbor until you can leave work* is a correct option. The test taker is demonstrating leadership in finding a solution to both the work dilemma as well as care for the child. In this instance, the test taker has found a solution which could extend for a period of time until the hospital situation may resolve. (+3 points)

C. *Call the school and have the child placed in after school care until you can get there* is a correct option. The test taker has leadership capabilities to recognize the need to remain at the hospital. However, after school care is time limited, so this is only a temporary fix for dealing with care of the child. (+ 3 points)

E. *Notify your supervisor of your child's situation and ask to leave as soon as possible* is another correct option. Here the test taker has recognized the need to take responsibility for patient care. However, in this scenario the test taker has not found an immediate solution for care of the children. (+3 points)

G. *Ask permission to speak to the supervisor's boss to discuss the need to leave the hospital is a neutral option*. Here the test taker recognizes the needs of the hospital, yet places the needs of family over the larger community. The test taker does recognize the need to notify the immediate supervisor before going above his/her head to a higher leader.(0 points)
F. *Stay at the hospital until such time as it is absolutely necessary to get your child* is an incorrect option. Here the test taker abandons the hospital when child care becomes critical. The test taker has not addressed the situation but is looking to avoid any conflict. (-1 point)
A. *Ask a colleague from the local area, who is not employed by the hospital, to come to the hospital to cover for you* is an incorrect option. While on the surface this would appear to handle both situations, it is a violation of patient confidentiality to ask an outsider to care for patients. In addition, the colleague has no legal responsibilities to the hospital and would be a liability issue were injury to occur. (-2 points)
B. *Inform your supervisor of your responsibilities for your child and leave the hospital* is an incorrect option. In this instance, the test taker has abandoned patients in the hospital and has demonstrated no ability to problem solve the scenario. This is conflict avoidance and demonstrates no leadership ability. (-3 points)

Question #5: Community Service
Toward the end of your day, a colleague from your unit tells you a patient who has chronic pain has been extremely rude to the team all day. This is not the first time this has occurred with this patient, however you have had good interactions with the patient during care. The incidences of rude behavior appear to be occurring more frequently. Your colleague seems very upset by this interaction.

Choose the THREE most appropriate responses.

○ A. Tell your colleague you will personally speak to the patient.

○ B. Go to the unit immediately and have a conversation with the patient.

○ C. Ask the patient to consider talking to a psychologist as everyone is trying to help.

○ D. Advise your colleague to ignore the patient as the pain is causing this behavior.

○ E. Encourage your colleague to apply to work in a different area of the hospital.

○ F. Recommend the team develop a plan about how to work with the patient.

○ G. Call the patient's family to discuss ways to work with this patient.

ANSWER: Correct FAB (Incorrect: CDEG (rank ordered))

This question deals with the test taker demonstrating an ability to participate in and be involved in the community. The test taker cares about the welfare of others.

F. *Recommend the team develop a plan about how to work with the* patient is a correct response. In this scenario, the test taker recognizes the larger community should work for a unified plan. This allows the patient's needs to be met while at the same time working to address the primary reason for this behavior which is abusive to staff. (+3 points)

A. *Tell your colleague you will personally speak to the patient is a correct option*. This demonstrates to the colleague you are listening to the issue, while at the same time giving the patient the opportunity to express their own opinion on the situation. The test taker recognizes there are two sides to every story, and because you have a good relationship with the patient you may be able to interact more effectively. (+3 points)

B. *Go to the unit immediately and have a conversation with the* patient is an appropriate response. The test taker recognizes verbal abuse toward staff should not be tolerated. In addition, the test taker will hear the patient's rationale for acting in the manner described, and because you have a good relationship with the patient you may be able to interact effectively. (+3 points)

C. *Ask the patient to consider talking to a psychologist as everyone is trying to help* is a neutral response. While this option does recognize fear and pain may be causing the patient to act out, the action requires the patient to take all actions. If the treatment team asked the patient to consider a consult for psychology, the

patient would still have the choice and the team would initiate the process. (0 points)

D. *Advise your colleague to ignore the patient as the pain is causing this behavior* is an incorrect response. Although the test taker is assuming pain is causing the behavior, the patient has not been asked and the response does not deal with the issue at hand. The patient has not been asked for the reasons, and the colleague is told to ignore the abuse. (-1 point)
E. *Encourage your colleague to apply to work in a different area of the hospital* is an incorrect response. This behavior does not address the problem and may only subject different team members to abuse. In addition, it forces an employee who may like their job to leave it due to inappropriate patient behavior. (-2 points)
G. *Call the patient's family to discuss ways to work with this patient* is an incorrect response. This response is a direct violation of patient confidentiality. Not only will this be a legal issue, it could be more harmful if the patient and family have additional issues toward each other. (-3 points)

Question #6: Communication

A 12-year-old patient is seeing you for a consult prior to surgery. The parents inform you they are Jehovah's Witnesses and will not allow the patient to have a blood transfusion if something should go wrong during the surgery.

Choose the THREE most appropriate responses.

○ A. Inform the surgical consultant in advance of the surgery the concerns brought up by the parents.

○ B. Tell the parents blood transfusions are unlikely during this surgery.

○ C. Consult with your supervisor about hospital guidelines for such events.

○ D. Ignore the parent's wishes because the child is a minor and is protected under law.

○ E. Explain to the parents you will seek additional guidance in this matter.

○ F. Encourage the parents to talk to the surgeon and express their concerns.

○ G. Listen to the parents and when appropriate continue to collect the information for your consultation.

ANSWER: AEC: Correct (Incorrect: FGBD (rank ordered)

This question is about respecting and communicating a patient's religious views in a manner which can best accommodate the religious views into appropriate care for the patient. The test taker demonstrates effective interpersonal and communication skills, and is able to identify a sense of caring about another individual's welfare The test taker should recognize a need to look for guidance to best negotiate this complicated scenario.

A. *Inform the surgical consultant in advance of the surgery the concerns brought up by the parents* is one of the most appropriate options. The test taker should recognize the hospital will need to be involved as the final decision maker in this scenario as it has legal, ethical, and cultural ramifications. (+3 points)
E. *Explain to the parents you will seek additional guidance in this matter* is another most appropriate option. This option recognizes the input from the parents and their cultural values, but also acknowledges such important decisions must be communicated to the larger hospital due to the ramifications which can accompany such a decision. (+ 3 points)
C. *Consult with your supervisor about hospital guidelines for such* events is another appropriate option. A test taker choosing this option recognizes the need to further their own learning, but does not recognize the greater hospital will need to be involved in the decision making. (+3 points)

D. *Encourage the parents to talk to the surgeon and express their concerns* is a neutral option. Here the test taker has heard the concerns of the parents but takes no action to assist them in the process. Instead the test-taker is relying on the parents to take the next step in the scenario. If the parents cannot reach the surgeon, have the concerns of the parents been adequately shared? (0 points)

E. *Listen to the parents and when appropriate continue to collect the information for your consultation* is not appropriate. In this option, the test taker does not even recognize or act on the parents' concerns. Here the test-taker identifies the most important thing to accomplish is to finish the consult. (-1 point)

F. *Tell the parents blood transfusions are unlikely during this surgery is* not an appropriate response. The test-taker should recognize the likelihood of the child needing a transfusion is not known, and to assume it is known would be lying to the parent. If the surgery were to proceed and the child needed blood, then the parents' decision has not been recorded and an inappropriate treatment could be provided. (-2 points)

G. *Ignore the parent's wishes because the child is a minor and is protected under law* is the least appropriate option. While the hospital can ultimately override a parents' decision regarding the care of a minor, this will cause significant conflict and is best avoided by sharing information prior to the surgery. (-3 points)

OPEN-ENDED STATIONS

The next six questions will be open-ended allowing you to write your own response. Please note spelling, grammar and other aspects of written communication will be considered in your response.

You will have a total of **45 minutes** to complete this section of the assessment tool.

You cannot navigate backward to see previous questions.

Scenario #1. After applying to the program of your choice, you are placed on a wait list. After waiting a few months, the school contacts you to let you know you were not accepted into the program. This is the only program you wanted to get into as it is close to where you live and you have always wanted to be in this profession.

What should you do?

_____

_____

_____

_____

_____

Scenario #2. During one of your clinical rotations as a student, you make a serious error while caring for a patient. The preceptor/clinical instructor gives you verbal feedback only and does not complete and official school evaluation of your performance. The preceptor/clinical instructor tell you he or she will not contact the program about the error.

What should you do?

_____

_____

_____

_____

_____

Scenario #3. You are escorting a patient to the area you will be providing care. As you travel through the facility, the patient begins to make racist, sexist, and ethnic remarks. You observe other patients and staff raising their eyebrows and glancing uncomfortably in your direction.

What should you do?

_____

_____

_____

_____

_____

Scenario #4. You are a student working in a busy facility with complex patients. One of your classmates is lazy, to the point of potentially compromising the care of patients at the facility. Staff from other departments have been making comments to you about how patients may be harmed by this classmate.

What should you do? Has anything in your background prepared you for such a situation?

_____

_____

_____

_____

_____

Scenario #5. You and another student both have clinical rotations at the same hospital, however, you do not share the same preceptor / clinical instructor. Today, your classmate arrives late, is in tears, and states an inability to continue to handle the stress of this clinical rotation. This is the third time in two weeks, your classmate has arrived late.

What should you do?

_____

_____

_____

_____

_____

Scenario #6. You are caring for a patient scheduled for heart surgery. The physician comes to the unit during the team meeting and informs the team the patient will most likely die, and completes the "Do Not Attempt Resuscitation" (DNAR) form. On your way to the patient's room, you observe the patient's family sitting in a waiting area just

down the hall from the meeting room. The family approaches you during your visit with the patient and asks you "so do you think my Mom is going to die?" It is clear to you the family overheard the team meeting.

What should you do?

_____

_____

_____

_____

_____

APPENDIX C: OPEN-ENDED STATIONS GRADING RUBRIC

Non-cognitive Attribute 1: Positive Self-Concept
This attribute assesses the student's ability to express confidence, strength of character, determination and independence.

The applicant:	Score 1 to 5	No evidence	Minimal evidence	Inconsistent evidence	Solid evidence	Outstanding evidence
language appears confident for future success.						
makes positive comments about self (specific comments about self – good learner, etc.)						
identifies a clear plan to re-apply or achieve a different goal.						
provides specific steps for how goal will be attained.						
describes future plans & experiences to enhance application						
acknowledges appropriate frustration and demonstrates resilience						
uses proper spelling and grammar.						

Non-cognitive Attribute 2: Realistic Self-Appraisal
This attribute assesses the applicant's ability to recognize and accept strengths and deficits, especially academic. The applicant works on self-development, applies critical thinking, and recognizes a need to broaden his/her individuality.

The applicant:	Score 1 to 5	No evidence	Minimal evidence	Inconsistent evidence	Solid evidence	Outstanding evidence
recognizes the error and acknowledges it with the school.						
asks for feedback about his/her strengths and/or weaknesses.						
recognizes the importance of feedback (positive or negative) on learning.						
discusses learning from the scenario.						
faces the problem with a determination to do better.						

acknowledges may have made a mistake (not fighting the system).		
uses proper spelling and grammar.		

Non-cognitive Attribute 3: Able to navigate system and culture The applicant exhibits a realistic view of the system based upon experiences, is committed to improving the system, and takes an assertive approach to dealing with wrongs. The applicant is not hostile to society.						
The applicant:	Score 1 to 5	No evidence	Minimal evidence	Inconsistent evidence	Solid evidence	Outstanding evidence
recognizes the unfairness of comments made by the patient.						
recognizes the need to address the situation.						
describes a resolution which minimizes continued comments by patient.						
is aware of impact of bias on the system (tries to maintain a professional environment).						
shows respect toward patient making comments despite comments (private area, polite, etc.)						
expresses ability to attempt to handle situation on own initially (does not go up chain of command at first)						
uses proper spelling and grammar.						

Non-cognitive Attribute 4: Leadership The applicant demonstrates leadership in any area of background (church, sport, non-educational groups).						
The applicant:	Score 1 to 5	No evidence	Minimal evidence	Inconsistent evidence	Solid evidence	Outstanding evidence
recognizes the need to address the situation.						
takes action and shows initiative by addressing the colleague and contacting appropriate parties.						
describes skills he/she has developed such as assertiveness.						

shows evidence of influencing others and being a good role model.		
is comfortable providing advice and direction to others.		
describes commitment (long-term) to skill development and responsibility for others.		
uses proper spelling / grammar.		

Non-cognitive Attribute 5: Community service
The applicant participates in and is involved in the community and cares about the welfare of others.

The applicant:	Score 1 to 5	No evidence	Minimal evidence	Inconsistent evidence	Solid evidence	Outstanding evidence
shows sustained (long-term) commitment to the care of his/her classmate.						
takes action and shows initiative by addressing with **colleague**.						
mentions the potential impact on patients at the facility.						
describes a plan to involve the community (school, site) in the care for the colleague.						
describes previous roles involving helping others outside of this scenario.						
promotes group problem solving (2 people working together) rather than solitary problem solving.						
uses proper spelling and grammar.						

Non-cognitive Attribute 6: Communication
The applicant demonstrates effective interpersonal and communication skills. The student is able to identify a sense of caring about another individual's welfare.

The applicant:	Score 1 to 5	No evidence	Minimal evidence	Inconsistent evidence	Solid evidence	Outstanding evidence
recognizes the need to provide the family information in a timely fashion.						
recognizes the information should be provided by an						

appropriate party (nurse, physician, etc.)		
recognizes communication should occur in a place of privacy (not in front of mother).		
demonstrates appropriate listening skills (look for words like dialogue, listens, etc.)		
recognizes this situation may be beyond their experience level.		
conveys empathy toward family (awareness of family feelings).		
uses proper spelling and grammar.		

REFERENCES

American Academy of Physician Assistants. (2012). *Competencies for the Physician Assistant Profession.* https://www.aapa.org/wp-content/uploads/2017/02/PA-Competencies-updated.pdf

American Association of Nurse Anesthetists. (2016). *Professional attributes of the nurse anesthetist: Practice considerations.* https://www.aana.com/docs/ default-source/practice-aana-com-web-documents-(all)/professional-attributes-of-the-nurse-anesthetist.pdf

American Physical Therapy Association. (2010). *Core values for the physical therapist and physical therapist assistant*. https://www.apta.org/siteassets/pdfs/policies/core-values-endorsement.pdf

Anderson, J., & Gerbing, D. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*(2), 155-173.

Artinian, N., Drees, B., Glazer, G., Harris, K., Kaufman, L., Lopez, N., & Michaels, J. (2017). Holistic admissions in the health professions: Strategies for Leaders. *College and University, 92*(2), 65-68.

Chana, T. (2017-2018). *Aggregate program data: 2017-2018 physical therapist education programs fact sheets.* Retrieved October 27, 2018, from http://www.capteonline.org/uploadedFiles/CAPTEorg/About_CAPTE/Resources/Aggregate_Program_Data/AggregateProgramData_PTPrograms.pdf

de Visser, M., Fluit, C., Cohen-Shotanus, J., & Laan, R. (2018). The effects of a non-cognitive versus cognitive admission procedure within cohorts in one medical

school. *Advances in Health Science Education, 23*(1), 187-200. doi:10.1007/s10459-017-9782-1

DiBaise, M., Salisbury, H., Hertelendy, A., & Muma, R. (2015). Strategies and perceived barriers to recruitment of underrepresented minority students in physician assistant programs. *Journal of Physician Assist Education, 26*(1), 19-27. doi:10.1097/jpa.0000000000000005

Dore, K., Kreuger, S., Ladhani, M., Rolfson, D., Kurtz, D., Kulasegaram, K., & Reiter, H. (2010). The reliability and acceptability of the multiple mini-interview as a selection instrument for postgraduate admissions. *Academic Medicine, 85*((10 Suppl)), S60-63. doi:10.1097/ACM.0b013e3181ed442b

Dore, K., Reiter, H., Eva, K., Krueger, S., Scriven, E., Siu, E., & Norman, G. (2009). Extending the interview to all medical school candidates - The computer-based multiple sample evaluation of non-cognitive skills (CMSENS). *Academic Medicine, 84*(10), S9-12.

Dore, K., Reiter, H., Kreguer, S., & Norman, G. (2017). CASPeR, an online pre-interview sceen for personal / professional characteristics: prediction of national licensing exam scores. *Advances in Health Science Education, 22*(2), 327-336. doi:10.1007/s10459-016-9739-9

Eva, K., Reiter, H., Rosenfeld, J., Trinh, K., Wood, T., & Norman, G. (2012). Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *JAMA, 308*(21), 2233-2240.

Eva, K., Rosenfeld, J., Reiter, & Norman, G. (2004). An admission OSCE: the mutliple mini-interview. *Medical Education, 38*(3), 314-326. doi:10.1046/j.1365-

2923.2004.01776.x

Field, A. (2018). *Discovering Statistics using IBM SPSS statistics* (5th ed.). SAGE Publications, Inc.

Follman, J., & Anderson, J. (1967). An investigation of the reliability of five procedures for grading English themes. *Research in the Teaching of English, 1*(2), 190-200. www.jstor.org/stable/40170454

Fraenkel, J., Wallen, N., & Hyun, H. (2014). *How to design and evaluate research in education* (9th ed.). McGraw-Hill Education.

Gould, W. (2014, July 1). *Improving diversity in graduate nurse anesthesia programs.* Retrieved October 16, 2018, from https://minoritynurse.com/improving-diversity-in-graduate-nurse-anesthesia-programs/

Grice, K. (2014). Use of multiple mini-interviews for occupational therapy admission. *Journal of Allied Health, 43*(1), 57-61.

Guyer, R., & Thompson, N.A., (2013). *User's Manual for Iteman 4.3.* Woodbury, MN: Assessment Systems Corporation.

Harvison, N. (2017-2018). *Academic Programs Annual Data Report.* Retrieved October 27, 2018, from https://www.aota.org/~/media/Corporate/Files/EducationCareers/Educators/2017-2018-Annual-Data-Report.pdf

Howe, K. (1997). *Understanding equal educational opportunities: Social justice, democracy, and schooling.* Teachers College Press.

Husbands, A., & Dowell, J. (2013). Predictive validity of the Dundee multiple mini-interview. *Medical Education, 47*, 717-725. doi:10.111/medu.12193

Jerant, A., Henderson, M., Griffin, E., Rainwater, J., Hall, T., Kelly, C., & Frank, P.

(2017). Reliability of multiple mini-interviews and traditional interviews within and betweeen institutions: A study of five california medical schools. *BMC Medical Education, 17*(190). doi:10.1186/s12909-017-1030-0

Kalsbeek, D. (2013). Employing noncognitive variables to improve admissions, and increase student diversity and retention. *Strategic Enrollment Management Quarterly, 1*(2), 132-152.

Kanny, E. (1993). Core values and attitudes of occupational therapy practice. *American Journal of Occupational Therapy, 47*(12), 1085-1086.

Kim, K., Nam, K., & Kwon, B. (2017, March). The utility of multiple mini-interviews: Experience of a medical school. *29*(1), 7-14. doi:dow.org/10.3946/kjme.2017.48

Koo, TK., Li, MY. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropractic Medicine*. 15(2), 155-163.

Meadows, J. (2014). *Blueprint for teaching cultural competence in physical therapy education.* Retrieved October 27, 2018, from https://www.apta.org/Educators /Curriculum/APTA/CulturalCompetence/

Oyler, D., Smith, K., Elson, E., Bush, H., & Cook, A. (2014). Incorporating multiple mini-interviews in the postgraduate year 1 pharmacy residency program selection process. *Am J Health-Syst Pharm,* 71(4), 297-304. doi:10.2146/ajhp130315

Patterson, F., Zibarras, L., & Ashworth, V. (2016). Situational judgement tests in medical education and training: Research, theory, ad practice: AMME Guide No. 100. *Medical Teacher, 38*(1), 3-17. doi:10.3109/0142159X.2015.1072619

Pau, A., Jeevaratnam, K., Chen, Y., Fall, A., Khoo, C., & Nadarajah, V. (2013). The multiple mini-interview (MMI) for student selection in health professions training -

a systematic review. *Medical Teacher, 35*(12), 1027-1041. doi:10.3109/0142159X.2013.829912

Portney, LG., Watkins, MP. (2020). *Foundations of clinical research: Applications to evidence-based practice* (4th ed.). F.A. Davis.

Rees, E., Hawarden, A., Dent, G., Hays, R., Bates, J., & Hassell, A. (2016). Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programs: A BEME systematic review: BEME Guide No. 3. *Medical Teacher, 35*(5), 443-455. doi:10.3109/0142159X.2016.1158799

Sawilowsky, S. (2000, April). Psychometrics versus datametrics: comment on Vacha-Haase's "reliabiity generalization" method and some EPM editorial policies. *Educational and Psychological Measurement, 60*(2), 157-173.

Sedlacek, W. (2017). *Measuring noncognitive variables: Improving admissions, success, and retention for underrepresented students* (1 ed.). Stylus Publishing, LLC.

Sedlacek, W., & Brooks, G. (1976). *Racism in American education: a model for change.* Nelson-Hall.

Shields, C. (2010). Transformative leadership: Working for equity in diverse contexts. *Educational Administration Quarterly, 46*(4), 558-589. doi:doi:10.1177/0013161x10375609

Shields, C. (2013). *Transformative leadership in education: Equitable change in an uncertain and complex world* (1st ed.). Eye on Education.

Shipper, E., Mazer, L., Merrell, S., Lin, D., Lau, J., & Melcher, M. (2017, July). Pilot evaluation of the computer-based assessment for sampling personal

characteristics test. *Journal of Surgical Research, 215*, 211-218.

Sternberg, R. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence* (1st ed.). Cambridge University Press.

Stratton, T., & Elam, C. (2014). A holistic review of the medical school admission process: examining correlates of academic performance. *Medical Education Online, 19*(1), 22919. doi:doi:10.3402/meo.v19.22919

Tracey, T., & Sedlacek, W. (1989). Factor structure of the non-cognitive questionnaire-revised across samples of black and white college students. *Educational and Psychological Measurement, 49*, 637-648.

van der Spuy, I., Busch, A., & Bidonde, J. (2016). Interviewers' experiences with two multiple mini-interview scoring methods used for admission into a master of physical therapy program. *Physiotherapy Canada, 68*(2), 179-185. doi:10.3138/ptc.2015-24E

Witzburg, R., & Sondheimer, H. (2013). Holistic review: Shaping the medical profession one applicant at a time. *New England Journal of Medicine, 368*(17), 1565-1567.

ABSTRACT

PILOT EVALUATION OF THE COMPUTER-BASED ASSESSMENT OF NON-COGNITIVE ATTRIBUTES OF HEALTH PROFESSIONALS (CANA-HP)

by

Sara F. Maher

December 2020

Advisor: Dr. Shlomo Sawilowsky

Major: Evaluation and Research

Degree: Doctor of Philosophy

To meet the needs of complex and/or underserved patient populations, health care professionals must possess diverse backgrounds, qualities, and skill sets. Holistic review has been used to diversify student admissions through examination of non-cognitive attributes of health care applicants. The objective of this study was to develop a novel methodology, the computer-based assessment of non-cognitive attributes of health professionals (CANA- HP), to effectively screen non-cognitive attributes of applicants. Three research questions were delineated; 1.) To determine the CANA-HP instrument reliability (internal consistency & interrater), 2.) To determine if the CANA-HP measured attributes of non-cognitive variables, as demonstrated by low construct validity scores when correlating the CANA-HP to traditional assessments reported to measure cognition, and 3.) To determine if differential item functioning on the CANA-HP revealed differences between groups based a variety of variables.

The study used a sample of convenience of students interviewed as part of the admission process into the occupational therapy program at Wayne State University

(N=37). Participants who consented to the study, completed a demographic survey followed by the 12 question CANA-HP. Data were analyzed using SPSS v. 25.0 (IBM, 2018) or Iteman v. 4.3 (ASC, 2013). Descriptive statistics of the sample population and 12 CANA-HP stations were computed. Cronbach's coefficient alpha was conducted on all of the stations for reliability, while interclass correlation estimates were run for interrater reliability. Pearson's correlation coefficients were calculated between CANA-HP scores and GRE / GPA scores at the time of program admission. Item difficulty, item discrimination, and bias were analyzed using mean average (*P*), *Rbpis,* and Fisher's exact tests respectively.

The six open-ended scenarios had minimally adequate internal reliability (α = 0.71), adequate item discrimination (*Rbpis* = 0.15 – 0.56), and adequate difficulty *(P* = 3.51 – 3.70). The traditional multiple choice questions need further refinement as these six scenarios had low reliability and discrimination. Initial results support the hypothesis of no correlation between the CANA-HP and standardized cognitive assessments (GRE and GPA scores). The one exception was non-science GPA which was significantly correlated to the total open-ended scores (*p* = .002) and total overall score (*p* = .008) and should be further examined. The CANA-HP is not biased toward the variables of sex, ethnicity, Pell grant status, family college history, or income level. Homogenous raters may improve interrater reliability which ranged from 0.67 – 0.91.

These results should be viewed with caution due to the small sample size conducted at only one university. Predictive validity of this methodology is needed. The CANA-HP remains a work in progress.

AUTOBIOGRAPHICAL STATEMENT

Sara F. Maher is a doctoral candidate in Education Evaluation and Research in the College of Education at Wayne State University in Detroit, Michigan. Her first degree was a Bachelor of Music Therapy from Western Michigan University. She worked in this field for a number of years, and made the decision to pursue a Master of Physical Therapy from Wayne State University. As part of her continuing education, she sought a Certificate in Orthopedic Manual Physical Therapy at Oakland University, followed by a DScPT degree. Her first faculty appointment was as an Assistant Professor at Oakland University in 2006. She was tenured and promoted to Associate Professor in 2012. In 2014, she graduated as a fellow of the Education Leadership Institute (ELI) of the American Physical Therapy Association, and was hired as Physical Therapy Program Director at Wayne State University. In 2016, she was promoted to Chair of the Department of Health Care Sciences, where she is blessed to work with students and faculty from six health care professions.

Her passion for innovative education has extended to volunteer work with the Federation of State Boards of Physical Therapy (Item Writer, Item Writing Coordinator, and Exam Development Co-chair), Foreign Commission for Credentialing of Physical Therapists (Board of Directors), Academy of Physical Therapy Education (Secretary), and American Council of Academic Physical Therapy (Education and Pedagogy Chair). She is currently the Item Writer Trainer for all physical therapists seeking to write multiple choice questions for the National Physical Therapy License Examination.