RESEARCH

Check for updates

Associations between an open-response situational judgment test and performance on OSCEs and fieldwork: implications for admissions decisions and matriculant diversity in an occupational therapy program

Mary Roduta Roberts[1]* , Cecilia Brito Alves[2] , Fu Chen[3] and Craig Richard St. Jean[4]

Abstract

Background Casper, an online open-response situational judgement test that assesses social intelligence and professionalism [1], is used in admissions to health professions programs.

Method This study (1) explored the incremental validity of Casper over grade point average (GPA) for predicting student performance on objective structured clinical examinations (OSCEs) and fieldwork placements within an occupational therapy program, (2) examined optimal weighting of Casper in GPA in admissions decisions using non-linear optimization and regression tree analysis to find the weights associated with the highest average competency scores, and (3) investigated whether Casper could be used to impact the diversity of a cohort selected for admission to the program.

Results Multiple regression analysis results indicate that Casper improves the prediction of OSCE and fieldwork performance over and above GPA (change in Adj. $R^2 = 3.2\%$). Non-linear optimization and regression tree analysis indicate the optimal weights of GPA and Casper for predicting performance across fieldwork placements are 0.16 and 0.84, respectively. Furthermore, the findings suggest that students with a slightly lower GPA (e.g., 3.5–3.6) could be successful in the program as assessed by fieldwork, which is considered to be the strongest indicator of success as an entry-level clinician. In terms of diversity, no statistically significant differences were found between those actually admitted and those who would have been admitted using Casper.

Conclusion These results constitute preliminary validity evidence supporting the integration of Casper into applicant selection in an occupational therapy graduate program.

Keywords Situational judgement tests, Program admissions, Occupational therapy, Fieldwork performance, OSCE

*Correspondence:
Mary Roduta Roberts
mroberts@ualberta.ca
[1]Department of Occupational Therapy, University of Alberta, Edmonton, AB, Canada
[2]Medical Council of Canada, Ottawa, ON, Canada
[3]Faculty of Education, University of Macau, Taipa, Macau, China
[4]Department of Communication Sciences and Disorders, University of Alberta, Edmonton, AB, Canada

Introduction and background

Strong academic and interpersonal skills are required to succeed as a health professional. The admission procedures of health professions programs typically use Grade Point Average (GPA) as a standard metric to evaluate an applicant's academic performance. In essence, GPA summarizes how well a student has performed academically over a certain period of time. Notably, GPA has consistently demonstrated statistically significant positive predictive relationships with future academic performance [2, 3]. Admissions procedures of health professions programs also look for evidence of strong non-academic skills (often called non-cognitive skills in previous literature) such as ethics, communication, judgement, critical thinking, and integrity [4, 5]. These non-academic skills are vital in creating and maintaining professional relationships with future colleagues and patients and are not strongly correlated with GPA performance [2, 5].

Occupational therapists (OTs), as regulated health professionals, are also expected to develop professional attributes such as empathy, communication, and ethical decision making [6]. However, assessing these non-academic skills remains a challenge through traditional OT program admission processes [7, 8]. Prioritization of academic achievement and standardized test scores for selection decisions can lead to issues of person-organization fit when new graduates move into the professional OT field [8, 9] and a higher risk of selecting applicants who perform well academically but have limited practical knowledge and skills for clinical practice [10, 11]. Current OT programs in Canada are, on average, 22–26 months in length [12]. Given this relatively short length of time to develop practice competencies, it is essential to select students with a baseline level of personal and professional characteristics from which continued development occurs in line with the profession's expectations [8].

Traditionally, assessment of non-academic skills has been accomplished using admission tools such as personal statements, references, or interviews [13]. However, screening and reviewing these admission tools can be resource-intensive, and the implementation of these tools often lacks reliability and validity, leading to biases in the admission process [4, 13, 14]. Alternatively, the use of multiple mini-interviews (MMIs) allows programs to broadly interview an applicant in multiple settings to gain a more accurate understanding of the applicant's non-academic abilities [4]. As described by Jerant and colleagues [15], some authors have concluded that MMIs have superior inter-rater reliability as compared with traditional interviews [11, 16] and that the approach yields moderate to high inter-rater reliability (range of Cronbach's alphas reported 0.65–0.98). However, the administration of MMIs is resource-intensive, thus making it

infeasible to be used as a non-academic skills screening tool for a high number of applicants [2].

One option currently being explored to effectively and efficiently screen applicants' non-academic skills is the situational judgement test (SJT). SJTs are applied in many occupations' selection processes as well as in admission procedures in medical schools [5]. This type of test requires applicants to respond to issues in hypothetical scenarios, which may be contextualized to a relevant setting or situation for which applicants apply [17]. In other words, SJTs examine what an applicant would do in a hypothetical situation.

Validity of situational judgement tests

Validity studies have been conducted in the context of using SJTs as part of admissions processes to health professions programs. Although not explicitly framed as such, the available research in the context of admissions and SJTs largely documents validity evidence of relationships between SJTs and other variables through predictive validity studies. SJTs have shown predictive validity in the context of medical education, with larger validity coefficients reported in graduate settings than in undergraduate settings (β=0.23) [18]. In a meta-analysis of students completing SJTs as part of undergraduate and graduate selection processes, 15 of 17 examined studies showed modest additional predictive power (typically between 5 −10%) over other selection assessments [18]. Further, Patterson and colleagues [14] reported an SJT demonstrated substantial incremental validity over application form questionnaire responses (ΔR^2=0.17) in predicting shortlisted candidates for postgraduate medical training selection. Lievens and Sackett [19] followed medical students for seven to nine years from admission to employment and found that scores on an interpersonal SJT at admission demonstrated incremental validity over academic measures in predicting fieldwork (ΔR^2=0.05) and job performance ratings (ΔR^2=0.05) (r=0.22 and 0.21, respectively). Since academic and non-academic skills are not mutually exclusive, researchers have identified an increasing need to investigate the incremental validity of non-academic assessments over conventional academic achievement metrics in medical education [11, 20, 21].

Casper

Casper–short for Computer-Based Assessment for Sampling Personal Characteristics–is an online open-response situational judgement test that assesses personal and professional characteristics, skills, and behaviors that inform social intelligence and professionalism [2]. Since Casper is an online test, it reduces the resource requirements for applicants and programs compared to MMIs, making it a much more accessible method to assess all

applicants' non-academic abilities [2]. Casper is scored by raters trained by the test administrators, making it more objective and less susceptible to bias than traditional application tools like interviews or reference letters [5]. Presently, Casper is used as part of admissions processes to health professions programs within North America and Australia.

Validity evidence for Casper also largely documents relationships between Casper and other variables in the form of correlational or predictive validity studies. In a study by Dore and colleagues [22], Casper scores were correlated against performance on both the Medical Council of Canada Qualifying Examination (MCCQE) Part I (end of medical school) and Part II (18 months into specialty training). Three to six years after entrance to medical school, the results showed a moderate predictive validity of Casper to national licensure outcomes of personal/professional traits with disattenuated correlations between 0.3 and 0.5. These results support the viability and reliability of a computer-based method for applicant screening in medical school admissions. Dore and colleagues [22] also found that Casper was moderately correlated with MMIs. Within the context of medical school admissions, Casper has been shown to predict medical students' future performance within a program and at postgraduate levels [2].

In addition to the prediction of future performance, there is evidence to suggest that a higher weighting of non-academic skills in the admission process may help to increase gender, racial, and ethnic diversity in health sciences programs [13, 23]. Findings from the literature suggest that medical students educated in diverse cohorts are better prepared to work with patients of variable circumstances and backgrounds [24]. As such, increasing diversity through admissions processes may be a necessary step to ensure graduating clinicians will be able to meet the changing needs of the communities and patients they serve. Some preliminary research on the use of Casper scores to increase the diversity of admitted applicants to their programs has demonstrated its potential with regard to gender, race, ethnicity, and socioeconomic status [13, 25], though other findings have been inconclusive [26, 27]. Such studies document preliminary consequential validity evidence of Casper's use during the admissions process for the purposes of increasing the diversity of admitted applicants.

Operational context for the study

The context for this study is admissions to a 26-month course-based entry-to-practice Master of Science in Occupational Therapy program in Western Canada. Applicants to the program are required to submit their transcripts from all post-secondary studies, their resume or curriculum vitae, two letters of reference, and must complete a statement of intent detailing their preparation, knowledge of, and suitability to practice occupational therapy. Applicants are rank ordered by GPA and approximately the top 300 from the applicant pool (approximately 40–50% depending on the year) are selected for a holistic file review conducted by members of the department. File review includes consideration of all their application materials. Files may be flagged for special consideration (i.e., a mature applicant with a lower GPA or completion of a research-based graduate degree) and are then reviewed by the admissions committee. Unless there are any major concerns noted during the file review, the applicant is recommended for admission. Offers of admission begin with the top-ranked candidate and continue down the list until all spaces within the program are filled.

Casper was newly introduced as an admissions requirement of the cohort reported in this study, but was not combined with GPA or other measures to inform selection decisions. For the occupational therapy program, Casper was viewed as potentially adding objective and reliable information about candidates' personal characteristics, which would be a valuable addition to the current admissions process. Additionally, the personal and professional characteristics, skills, and behaviors measured by Casper, such as communication, empathy, problem-solving, professionalism, and self-awareness, align with the competencies assessed within the program through performance-based activities such as objective structured clinical examinations (OSCE) and fieldwork placements. As such, Casper scores could assist the program in identifying applicants with these valued personal competencies and high potential for success as a future health professional.

Figure 1 is a graphical representation of the Master of Science in Occupational Therapy curriculum where courses are organized according to six themes. This study focuses on assessments of practice-based activities within the 'Integration and Application' and 'Fieldwork' curriculum themes which include OSCEs and fieldwork placements. Both themes emphasize the application and demonstration of professional competencies, including non-academic skills learned throughout the program.

Theoretical framework - validation

We position our study within contemporary notions of assessment validation [28, 29]. According to Kane [28], validation entails two steps. First is the articulation of key claims in a proposed interpretation or use of the assessment (i.e., interpretation-use argument; IUA), which includes identifying associated assumptions and inferences. Second is developing a plan to test and evaluate the stated claims, assumptions, and inferences. This plan, known as the validity argument, informs the collection of
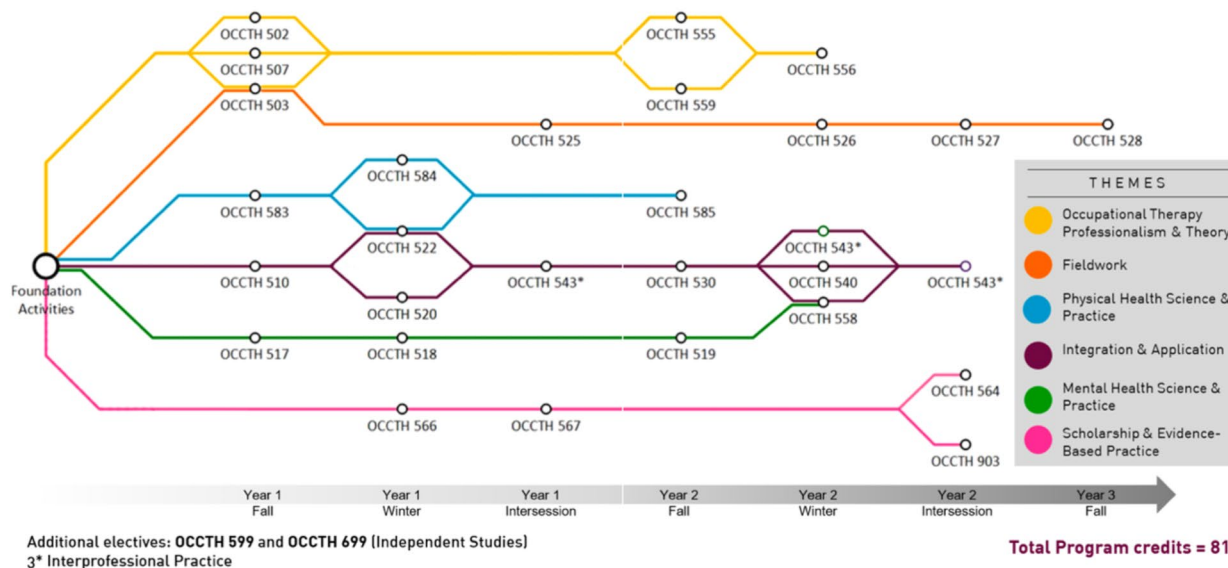
Fig. 1 Curriculum map of the course-based master's program in occupational therapy

empirical evidence of which its synthesis and appraisal provides support for or against the proposed interpretation and use of the assessment. A validity argument can be structured or prioritized around the four stages of inference moving from observation to interpretation and decision: scoring, generalization, extrapolation, and implication. These stages of inference within the validity argument direct the collection of different sources of validity evidence, which include those specified by Messick [30]: content, response process, internal structure, relationships with other variables, and consequences.

Following the validation framework described by Kane [28], we evaluated evidence to support the following central claim of proposed use: *As a measure of personal and professional competencies that inform social intelligence and professionalism, Casper can be used to inform admissions decisions to select a diverse cohort of individuals with a high potential to succeed in practice-based activities requiring non-academic skills.* Within this central claim, we investigate and prioritize two types of inferences—extrapolation and implication. We focus our attention on these two inferences because the test developer must undertake the evaluation of claims related to the scoring and generalization inferences prior to administration. Strong evidence in support of the scoring and generalization inferences has been established and documented in a technical manual by the test developer [1]. However, collecting evidence to evaluate claims related to the extrapolation and implication inferences relies on student data from programs using Casper as part of their admissions process.

For each of the extrapolation and implication inferences, our validity argument entailed articulation of key assumptions and the evidence gathered to evaluate each inference. Relevant to the extrapolation inference is the assumption that the Casper score indicates an accurate level of the student's personal and professional characteristics measured that is consistent with how those characteristics are demonstrated in their performance of practice-based activities requiring non-academic skills. The evidence to evaluate this assumption was in the form of observed associations between Casper and OSCEs and fieldwork performance. Relevant to the implication inference is the assumption that including Casper as part of the admissions process supports the selection of a more diverse cohort of students when compared to selection using GPA alone. The evidence to evaluate this assumption was the extent of demographic group differences as measured by the percentage overlap of membership between the two groups selected using different admissions metrics.

Study objectives

Following from Kane's validation framework [28], the focus of this study is evaluating the validity argument and appraising the collective evidence so as to come to a conclusion on whether it supports the proposed use and its consequences as outlined in the IUA. Therefore, the objectives of this study are to gather and evaluate validity evidence regarding relationships between Casper and other variables and consequences. Sources of validity evidence will be in the form of (1) incremental validity

of Casper over GPA for predicting performance on the best available in-program measures that require the use of non-academic skills, i.e., objective structured clinical examinations (OSCEs) and fieldwork (Competency-Based Fieldwork Evaluation), (2) associations between different Casper and GPA weightings with non-academic performance, and (3) effects of differentially weighting Casper and GPA on the diversity (i.e., demographic composition) of applicants selected for admission.

Method

Data sources

Applicant data (*n*=548) from one admissions cycle to a 26-month course-based Master's program in occupational therapy was obtained. Applicant data included demographics, degree program, GPA, and Casper scores. For the 125 students admitted from this applicant pool, scores from four OSCEs (i.e., OSCE 510, 520, 530, and 540) and ratings from four fieldwork placements (i.e., Placements 1–4, denoted P_1 to P_4) were added to the data set.

Instruments

Casper

The Casper test [1] is composed of 12 sections. Each section contains either a brief video or word-based scenario, followed by a series of three open-ended questions which participants have five minutes to answer. Each scenario relates to one or more personal characteristics. Each section is scored by a different set of raters trained by the test administrator Acuity Insights, where collectively, the rater pool is intended to reflect the diversity of the population. The 12 section scores are averaged to generate the overall score, which is reported as a *z*-score.

GPA

Admission GPA was calculated over the last 60-credits, or two full-time equivalent years of applicants' coursework and used in the analysis with a scale of 0.0 to 4.0.

OSCE

The four OSCEs discussed in this paper, OSCEs 510, 520, 530, and 540, occur within the Integration and Application theme at the end of each of the four semesters of coursework (i.e., OCCTH 510, 520, 530, and 540, as displayed in Fig. 1). No further OSCEs are administered in the program. The first OSCE (OSCE 510) is established in the context of physical/musculoskeletal assessment, where students complete a basic client interview that is structured using an occupation-based model. This OSCE has a particular emphasis on demonstration of interactive reasoning, which is the application and monitoring of interpersonal skills to build and sustain a therapeutic interaction. The second OSCE (OSCE 520) is established

in the context of a simulated case review with a fieldwork supervisor, where the client presents with both physical and mental health issues. This OSCE examines communication, professionalism, and clinical reasoning involved in the identification and justification of occupational performance issues and intervention strategies. In the demonstration of effective clinical reasoning, students must also display appropriate knowledge of the client, relevant theories and models, and the evidence base supporting an intervention strategy.

The third OSCE (OSCE 530) is established in the context of pediatrics, family-centered care, and neurological rehabilitation, where students engage in a simulated rural home visit with a mother to discuss intervention strategies for her 18-month-old child who has cerebral palsy. The final OSCE (OSCE 540) is established across multiple contexts that students have been exposed to throughout the program (e.g., older adults, community-based practice). Students randomly select a case write-up and complete a consultation with a standardized patient which includes assessment, communication of assessment findings, and recommendations for intervention and follow-up. Both OSCE 530 and 540 are designed to elicit demonstration of the same competencies as OSCE 520, but the client scenarios increase in complexity.

For each OSCE, a rubric was used to assess student performance. For OSCE 510, students were assessed on four domains: (1) Professionalism, (2) Communication, (3) Knowledge of the client, and (4) Clinical reasoning. For OSCE 520–540, students were assessed on the same four domains, plus (5) Theory, models, frames of reference, and (6) Evidence-based practice. Assessors provided a holistic score between 0 and 50 for OSCE 510, 520, and 540, and between 0 and 100 for OSCE 530. All OSCE scores were standardized to facilitate comparability.

Fieldwork

The four fieldwork placements discussed in this paper, OCCTH 525, 526, 527, and 528, which henceforth will be referred to as *Placements 1, 2, 3, and 4*, occur outside the semesters of coursework. Placement 1 occurs after the first year, or two semesters, of coursework. Placement 2 occurs after the third semester of coursework, and Placements 3 and 4 occur after all coursework is completed. Students are required to complete placements in three broad occupational therapy practice areas, in any order, by the end of their program. The practice areas are physical medicine (e.g., orthopedics, general medicine), mental health (e.g., psychiatry inpatient and outpatient), and community/rural (e.g., schools, home care, rural). Of note is that placements can span more than one practice area but are classified according to the predominant services provided. Additionally, the clients seen within each of these practice areas can range across the lifespan.

Fieldwork placements can embody practice in specific contexts (e.g., neurorehabilitation) and with specific populations (e.g., older adults). Evaluation of fieldwork performance focuses on the student's demonstration of practice and non-academic competencies such as clinical reasoning, establishing therapeutic and collaborative relationships, adhering to ethical practice standards, and communication.

In each of the fieldwork placements, students are evaluated using the Competency-Based Fieldwork Evaluation tool (CBFE) [31], which has seven core competencies of practice (CBFE 1 to 7). See Table 1 for definitions for each competency as well as scoring instructions.

Data analyses
Objective 1
Associations between Casper scores, GPA, four OSCE scores and four CBFE competency scores were examined using Spearman's correlation. We correlated GPA with four OSCE scores and twenty-eight fieldwork competencies, and conducted similar analyses for Casper. The incremental validity of Casper over GPA was

Table 1 Definitions of the competency-based Fieldwork evaluation (CBFE) competencies

Competencies	Definition
Practice knowledge	Discipline-specific theory and technical knowledge
Clinical reasoning	Analytical and conceptual thinking, judgement, decision making and problem-solving
Facilitating change with a practice process	Assessment, intervention planning, intervention delivery and discharge planning
Professional interactions and responsibility	Relationships with clients and colleagues and legal and ethical standards
Communication	Verbal, non-verbal and written communication
Professional development	Commitment to profession, self-directed learning and accountability
Performance management	Time and resource management, and leadership
Overall performance	Holistic rating taking into consideration performance across the seven competencies. Each of the seven competencies and overall performance (CBFE 8) are rated on a scale from 1–8, indicating the student's progress from entry-level student to entry-level clinician. In extraordinary situations, a rating of U (Unacceptable) or E (Exceptional) can be used
Scoring the CBFE	A range of scores within the 1–8 corresponds to expectations along with a developmental progression in fieldwork. For example, Level 1 placement (i.e., Placement 1) expectations align with Stage 1 competencies, where student scores may range from 1–3. A score of 3 indicates mastery of Stage 1 competencies and a transition to Stage 2, with scores ranging from 3–6. This scoring logic can be extended to Stage 3 competencies

evaluated using hierarchical multiple regression, examining changes in R-squared values as well as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These criteria provide a quantitative measure of the trade-off between model fit and complexity. Model 1 considered GPA as the predictor variable, and Model 2 included Casper as an additional predictor. The smaller the AIC/BIC, the better the model. Additionally, Cohen's d analyses were performed in this study to measure the effect size, providing deeper insight into the magnitude of the observed differences.

Objective 2
Non-linear optimization [32] and regression tree analyses [33] were conducted to identify the optimal weights for GPA and Casper in admitting students with high future fieldwork placement performance, representing a novel analytical approach within this body of literature. An optimization problem is to find the extrema of an objective function subject to a set of constraints. Non-linear optimization represents a process used to identify the best solution or outcome for a problem where the involved variables are not linearly associated with each other [32]. In simple words, this process is analogous to finding non-straight paths leading to the highest peak or the lowest valley on a landscape that meets a few pre-established conditions. In our study, five admission metrics were selected as variables for solving the non-linear optimization problem: GPA, degree, Casper, Indigenous status, and age. Specifically, the objective function of non-linear optimization is defined to maximize the correlation between the weighted admission scores and students' future fieldwork placement performance. As students' undergraduate degree program is a categorical variable with five major categories (General science, General arts, Health, Kinesiology, and Psychology), it was converted to four dummy code variables (D1 to D4, representing comparisons between degrees) by the dummy coding method for non-linear optimization. As such, the weighted admission scores are given by:

Weighed Score $= w_1 \times$ GPA $+ w_2 \times$ Casper $+ w_3 \times$ Indigenous $+ w_4 \times$ Age $+ w_5 \times$ D1 $+ w_6 \times$ D2 $+ w_7 \times D_3 + w_8 \times$ D4. In addition, all the admission metrics should satisfy the following two constraints: $w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8 \geq 0$ and $w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 = 1$. Having estimated the weights for the five admission metrics by non-linear optimization, we can compare the relative strength of each admission metric in association with students' future fieldwork placement performance.

To cross-validate the results by non-linear optimization and further examine the weights of the five admission metrics, we employed regression tree analysis, a predictive modeling approach widely used in data mining and machine learning. In simple words, regression

tree analysis predicts the value of a target by splitting the data into smaller and smaller portions according to certain conditions. This process can be represented by a tree structure. Suppose a tree is established to predict a student's future academic performance based on a few student features such as GPA. Each branch of the tree represents a decision question, such as "Is his or her GPA above 3.5?". The question classifies the student into a smaller group of the training data. As the branches go down, the student will end up in a "leaf" that makes the final prediction based on the data associated with the leaf group. It is similar to human problem-solving, which is breaking down a task into simpler and simpler steps. More specifically, regression trees predict the values of a target variable based on a set of features by iteratively splitting the data into sample regions [33]. To make a prediction for a given sample, regression trees classify the sample into a region based on the values of its features. Typically, the mean of all training samples in the region to which the sample belongs is used as a predicted value. To build a regression tree, the model recursively makes splits on the tree nodes (a node can be considered a subgroup of samples after splits) using different features. As such, the selection of features for each split determines the effectiveness of a regression tree. To choose the best-performing feature to partition samples for a split, the model uses selection criteria such as variance and the Gini index, which measures the homogeneity of nodes before and after splits. For a given split, the feature resulting in the greatest reduction of the selection criterion is used to partition the samples. Across all splits in a tree for which a feature was used, the total reduction of the criterion brought by the feature can be used to quantify the importance of the feature. This study used the feature importance scores from the regression tree as the weights for each admission metric.

We conducted two rounds of regression tree analyses. In the first round, the features used to build regression trees were the five admission metrics used in the non-linear optimization. The first round of regression tree analysis aimed to examine whether GPA and Casper were more influential than the other three admission metrics in predicting students' future fieldwork placement performance. In the second round, we only used GPA and Casper scores for regression tree analyses, and their weights were averaged across regression trees associated with different fieldwork placements to obtain final weights to calculate students' admission scores (i.e., $w_1 \times$ GPA $+ w_2 \times$ Casper; both were standardized). The historical admission data used in this study indicated that 179 students were offered admission (labeled as 'accepted,' 'accepted but declined,' or 'admitted' in the applicant pool). To reflect the actual admission outcomes, we selected the top 179 students with the highest weighted

scores as the weighting-selected students. The results from non-linear optimization and regression tree analysis can be used to cross-validate each other. Additionally, we conducted *t*-tests to determine whether GPA, Casper score, and age changed using weighted admission scores.

Previous regression tree analyses were performed to determine the optimal weights for GPA and Casper. To further validate the effectiveness of Casper in admissions, this study also aimed to investigate whether students admitted by different weighting scenarios of GPA and Casper show different future academic performance. Therefore, we conducted additional weighting scenario analyses to provide additional evidence from the perspective of high-performing students. Specifically, we investigated whether reducing Casper's weight in admissions correlates with lower fieldwork placement performance among admitted students. To achieve this, a composite admissions score was calculated based on GPA and Casper for a series of weighting scenarios (i.e., GPA was assigned a weight ranging from 0.1 to 0.9). For example, given weights of 0.1 and 0.9 for GPA and Casper, respectively, the admission scores can be calculated as $0.1 \times$ GPA $+ 0.9 \times$ Casper (both were standardized). To evaluate the effectiveness of each weighting scenario, we assessed whether the top students identified by each weighting system were truly high performing in their future studies. For each scenario, we selected the top 30 admitted students in terms of the weighted scores and calculated their average competency scores for each fieldwork placement. Consistent with the commonly adopted definition of top performers in the literature, we focused on the top 25% of admitted students, corresponding to the top 30. Additionally, a sample size of 30 is commonly used in statistics to provide reliable estimates of population parameters [34].

Objective 3
To examine the effect of differentially weighting Casper and GPA on the diversity (i.e., demographic composition) of applicants selected for admission, we conducted chi-square tests to compare demographic and degree characteristics between the actually admitted students and the set of students that would have been selected had Objective 2's optimized weighting been used in the admissions process. The optimal weights of GPA and Casper were obtained from the second-round regression tree analysis.

In summary, this study will present four types of analyses: (1) Hierarchical multiple regression analyses, (2) non-linear optimization, (3) regression tree analyses, and (4) chi-square analysis, which will be detailed in the upcoming results section.

Table 2 Descriptive statistics for GPA, Casper, OSCEs and fieldwork placements

	GPA	Casper	OSCE 510	OSCE 520	OSCE 530	OSCE 540	Placement 1*	Placement 2*	Placement 3*	Placement 4*
Mean	3.8	0.3	0.0	0.0	0.0	0.0	3.5–3.8	5.8–6.3	7.1–7.6	8.0–8.2
Std Dev	0.1	0.8	1.0	1.0	1.0	1.0	0.9–1.2	0.7–1.0	0.7–0.8	0.5–0.7
Minimum	3.3	-1.8	-2.5	-1.9	-2.2	-2.7	1.5–2.0	3.0–3.5	5.0–6.0	4.0–5.0
Maximum	4.0	2.0	2.0	2.7	1.9	2.0	7.0–9.0	9.0	9.0	9.0
N	120	120	121	119	119	115**	120	118	117	111

Note. *Students receive eight scores for each fieldwork placement (i.e., CBFE 1–8). Due to space limitations, only the minimum and maximum mean values for each CBFE competency are provided

** Two students did not complete all the course activities and were excluded from this table

Table 3 Statistically significant correlations of GPA and Casper with fieldwork competencies and OSCE scores

Variable	Placement/OSCE	Correlation Coefficient	*p*-value
Casper	P2_2	0.279	0.004
Casper	P2_3	0.196	0.043
Casper	P2_4	0.213	0.027
Casper	P2_5	0.245	0.011
Casper	P2_6	0.248	0.010
Casper	P3_1	0.280	0.003
Casper	P3_2	0.259	0.007
Casper	P3_3	0.209	0.030
Casper	P3_4	0.191	0.048
Casper	P3_5	0.262	0.006
Casper	P3_7	0.193	0.045
Casper	P3_8	0.293	0.002
Casper	P4_2	-0.206	0.032
Casper	OSCE 510	0.192	0.046
Casper	OSCE 540	0.264	0.006
GPA	Neither the Placements nor the OSCEs showed a statistically significant correlation ($p < 0.05$)		

Notes. 1. Significance at p-value < 0.05

2. Two students did not complete all the course activities for 540 and were excluded from this table

Results
Descriptives
Descriptive statistics for GPA, Casper scores, OSCEs, and fieldwork scores are summarized in Table 2. The high GPA mean (3.8) and low standard deviation (0.1) indicate that candidates with high academic performance were selected for the program. Greater variability was observed with Casper scores ($M=0.3$, $SD=0.8$), where the mean±2 *SD* approached the limits of the range of scores. The ranges of the placement averages suggest that, as expected, the higher the placement (e.g., Placement 4), the higher the range of the scoring average.

Objective 1
Associations between Casper scores, GPA, OSCE and fieldwork scores
Correlations were statistically significant between Casper scores and two OSCEs (510 and 540), five competencies

on the second placement, seven competencies of the third placement, and one competency on the fourth placement. No correlation was statistically significant for GPA. Table 3 displays the statistically significant correlations between each measure ($p < 0.05$).

Prediction of student performance using hierarchical multiple regression
Table 4 shows the regression coefficients for Model 2, which includes both GPA and Casper scores. Results show that Casper has a statistically significant positive effect on OSCE 510, OSCE 540, two competencies on the first placement, five competencies on the second placement and four competencies on the third placement, indicating that as Casper scores increase, the scores in those courses are likely to increase as well. Only one parameter estimate for GPA (P2_5) was statistically significant ($p < 0.05$); however, it was in the opposite direction of what we would expect (the higher the GPA, the lower the score on P2_5).

Table 5 shows two models: Model 1, which solely incorporates GPA, and Model 2, which incorporates both GPA and Casper scores. The results reveal that Model 2 exhibited higher average R^2 values and lower AIC and BIC values, indicating a superior fit compared to Model 1. Adding Casper scores improved the overall model's R^2 value for all four OSCEs and all CBFE competencies. On average, incorporating Casper scores improved performance prediction beyond GPA alone (change in $R^2 = 3.2\%$).

As shown in Fig. 2, the inclusion of Casper scores improved the overall R^2 value of the model for all four OSCEs and CBFE competencies. This suggests that the inclusion of Casper outperformed the model that solely relied on GPA, showcasing its superior predictive performance across most of the evaluated competencies.

Cohen's *d* was computed to measure the effect size of GPA and Casper scores on the evaluated competencies. Cohen's *d* measures effect size, indicating the standardized difference between two means. Based on benchmarks proposed by Cohen [35], Cohen's *d* typically categorizes effect sizes as small ($d=0.2$), medium ($d=0.5$), and large ($d=0.8$). Table 6 shows that Casper's

**Table 4** Regression analysis of OSCEs and placement scores: effects of GPA and Casper

OSCE/ Placement	Predictor	DF	Unstandardized Coefficient (B)	Standardized Coefficient (Beta)	Std. Error	t-Value	*p*-value
OSCE 510	Intercept	1	-4.08	0.00	2.58	-1.58	0.12
	GPA	1	1.07	0.14	0.69	1.55	0.12
	Casper	1	0.25	0.20	0.11	2.18	**0.03**
OSCE 540	Intercept	1	-3.10	0.00	2.53	-1.23	0.22
	GPA	1	0.80	0.11	0.67	1.18	0.24
	Casper	1	0.41	0.32	0.11	3.61	**0.00**
P1_5	Intercept	1	0.53	0.00	2.61	0.20	0.84
	GPA	1	-0.16	-0.02	0.70	-0.23	0.82
	Casper	1	0.27	0.22	0.11	2.39	**0.02**
P1_7	Intercept	1	-0.36	0.00	2.62	-0.14	0.89
	GPA	1	0.08	0.01	0.70	0.11	0.91
	Casper	1	0.24	0.19	0.11	2.05	**0.04**
P2_2	Intercept	1	3.38	0.00	2.55	1.32	0.19
	GPA	1	-0.92	-0.12	0.68	-1.36	0.18
	Casper	1	0.33	0.26	0.11	2.87	**0.00**
P2_3	Intercept	1	2.79	0.00	2.63	1.06	0.29
	GPA	1	-0.76	-0.10	0.70	-1.09	0.28
	Casper	1	0.24	0.19	0.12	2.08	**0.04**
P2_4	Intercept	1	3.26	0.00	2.59	1.26	0.21
	GPA	1	-0.90	-0.12	0.69	-1.30	0.20
	Casper	1	0.35	0.27	0.12	3.03	**0.00**
P2_5	Intercept	1	7.96	0.00	2.51	3.16	0.00
	GPA	1	-2.15	-0.28	0.67	-3.20	**0.00**
	Casper	1	0.34	0.27	0.11	3.07	**0.00**
P2_6	Intercept	1	4.44	0.00	2.55	1.74	0.09
	GPA	1	-1.21	-0.16	0.68	-1.78	0.08
	Casper	1	0.39	0.30	0.11	3.40	**0.00**
P2_7	Intercept	1	1.60	0.00	2.53	0.63	0.53
	GPA	1	-0.44	-0.06	0.67	-0.65	0.52
	Casper	1	0.23	0.19	0.11	2.03	**0.05**
P2_8	Intercept	1	4.51	0.00	2.56	1.76	0.08
	GPA	1	-1.23	-0.16	0.68	-1.79	0.08
	Casper	1	0.32	0.26	0.11	2.82	**0.01**
P3_1	Intercept	1	0.71	0.00	2.64	0.27	0.79
	GPA	1	-0.21	-0.03	0.70	-0.30	0.77
	Casper	1	0.24	0.19	0.12	2.04	**0.04**
P3_2	Intercept	1	2.78	0.00	2.61	1.06	0.29
	GPA	1	-0.76	-0.10	0.70	-1.10	0.28
	Casper	1	0.29	0.23	0.12	2.45	**0.02**
P3_5	Intercept	1	2.04	0.00	2.62	0.78	0.44
	GPA	1	-0.56	-0.07	0.70	-0.80	0.42
	Casper	1	0.26	0.20	0.12	2.19	**0.03**
P3_8	Intercept	1	2.78	0.00	2.60	1.07	0.29
	GPA	1	-0.76	-0.10	0.69	-1.10	0.27
	Casper	1	0.30	0.24	0.12	2.57	**0.01**

Notes. 1. The estimates in this table are based on Model 2 (Y = b0 + b1(GPA) + b2(Casper))

2. Two students did not complete all the course activities for 540 and were excluded from this table

Table 5 Model evaluation results using average R^2, AIC, and BIC values

Number in Model	Variables in Model	*R*-Square	AIC	BIC
1	GPA	0.006	2.263	4.234
2	GPA + Casper	0.038	0.375	2.533
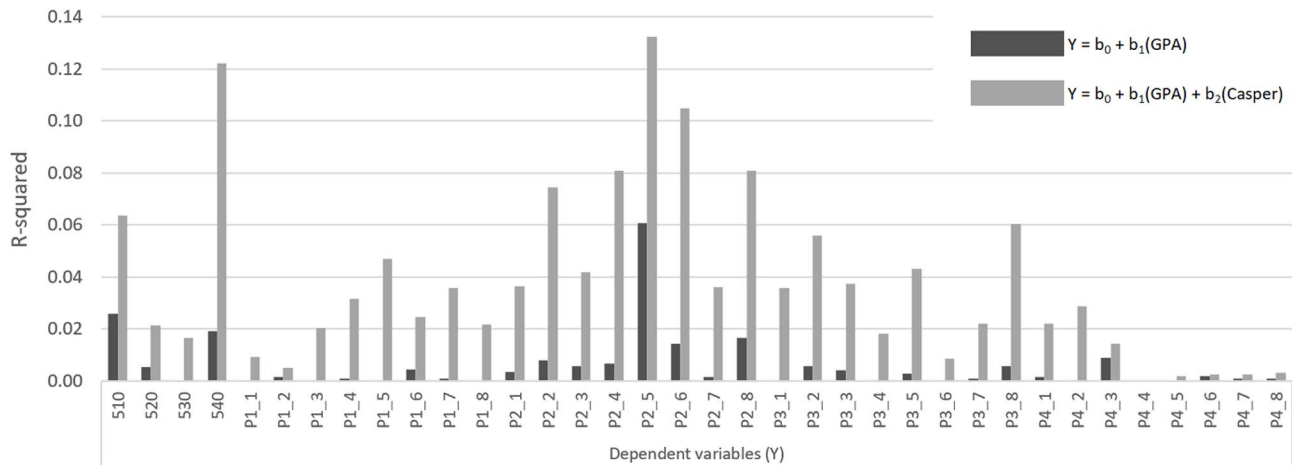
Fig. 2 R-squared values for OSCE and placements by model 1(GPA) and model 2 (GPA + Casper). *Note.* Two students did not complete all the course activities for 540 and were excluded from this graph

Table 6 Average Cohen's *d* and percent of competencies by effect size ranges

	Average Cohen's d	Cohen's d (%)			
		[-∞, 0.20)	[0.2 to 0.5)	[0.50, 0.80)	[0.80, +∞)
GPA	-0.07	91.7	8.3	0.0	0.0
Casper	0.26	30.6	52.8	16.7	0.0

average Cohen's *d* values are larger than GPA's. A negative value of Cohen's *d* suggests that the means of the GPA and some competencies are in opposite directions. In fact, when considering the GPA, Cohen's *d* in roughly 92% of the evaluated competencies is smaller than 0.20, indicating a minimal effect. In contrast, only 30.6% of the competencies had Cohen's *d* values lower than 0.20 when Casper is considered.

Objective 2

Weights of admission metrics identified by non-linear optimization and regression tree analysis

The optimal weights of Casper identified by non-linear optimization are much higher than other admission metrics, and GPA is less associated with students' competency scores (see Table S1 in the electronic supplementary material). In addition, regarding the weights identified by regression tree analysis, the first set of regression tree analyses was conducted with the same five admission metrics as the non-linear optimization analysis. Figures 3, 4, 5 and 6 show that Casper has a much higher importance weight, followed by GPA, for predicting students' performance across each fieldwork placement (e.g., 70% for Casper and 10% for GPA in the third fieldwork placement).

The second set of regression tree analyses was conducted with GPA and Casper as the only predictors. Table 7 shows the importance scores for each fieldwork

placement. Overall, Casper is a dominant admission metric in predicting students' competency scores in fieldwork placements. The average weights of GPA and Casper across all regression tree analyses are 0.16 and 0.84, respectively.

Comparison of fieldwork placement performance between different weighting scenarios

Generally, except for the fourth placement, increasing GPA weight and decreasing Casper weight was associated with lower average competency scores (see Figures S1 to S4 in the electronic supplementary material). GPA might not be a significant predictor of top-performing students' fieldwork performance, though this trend is not salient for the fourth placement. However, this finding was expected because the variances of competency scores in the fourth placement are extremely low (see Table 2).

Objective 3

Demographic differences between admitted students and students selected by weighted admission scores

Tables 8 and 9 display the admission metric and demographic differences, respectively, between admitted students and students selected for admission by weighted admission scores, which were calculated based on the optimal weights of GPA and Casper (i.e., 0.16 and 0.84, respectively) from the second set of regression tree analyses. The actual admitted students showed significantly higher GPA scores ($t=11.57$, $p<0.001$) but significantly lower Casper scores ($t = -21.06$, $p<0.001$) than the students selected by weighted admission scores. The age difference between the two groups was not significant ($t=0.05$, $p=0.96$). Chi-square tests showed that the two groups did not differ from each other in terms of number of Indigenous students, female students, and degree
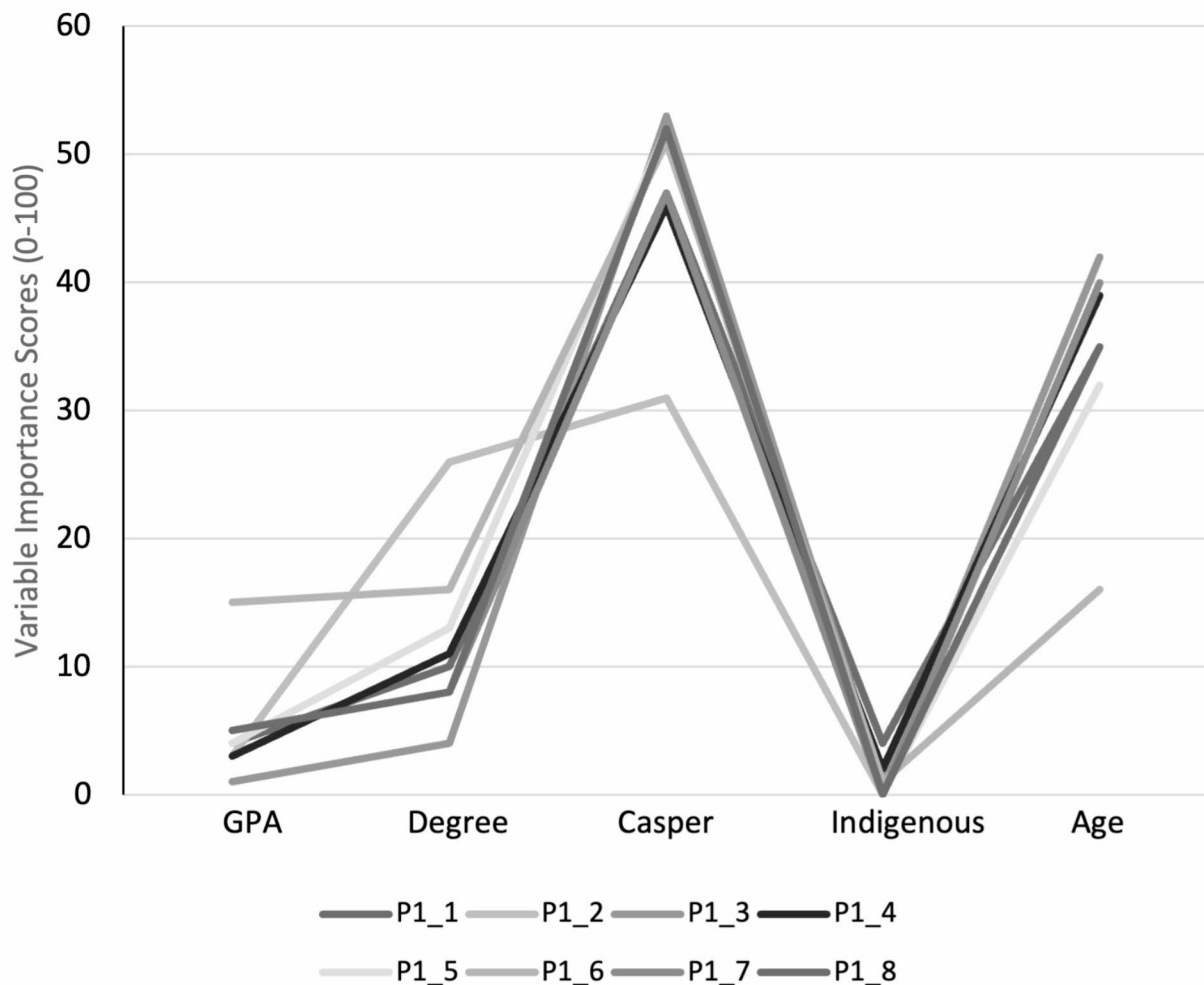
Fig. 3 Variable importance of each admission metric for the competency scores in the first fieldwork placement

distribution ($X^2 = 0.08$, $df = 1$, $p = 0.77$, $X^2 = 0.10$, $df = 1$, $p = 0.75$, and $X^2 = 2640.9$, $df = 2714$, $p = 0.84$, respectively). The percentage of overlapping students between the two groups is 46.37%.

Discussion

Presently, Casper is being used as part of a holistic file review during the admissions process within our occupational therapy program. To support the inclusion of Casper as an admissions requirement, we investigated whether Casper scores are associated with future performance in practice-based activities that require non-academic competencies within our program, such as objective structured clinical examinations (OSCEs) and fieldwork. Additionally, we were interested in the potential for Casper to increase the diversity of applicants offered admission to the program. Therefore, the purpose of this study was to gather and evaluate validity evidence for the use of Casper as part of the admissions process

within an occupational therapy program. Overall, results from the three analytic approaches show that Casper is a better predictor than GPA for predicting student performance on OSCEs and most fieldwork—assessments that require the use of non-academic competencies such as communication and professionalism.

When analyzing the associations between Casper scores, GPA, OSCE and fieldwork scores, results indicated that the correlations were statistically significant between Casper scores and two OSCEs, on multiple competencies on the fieldwork placements. No correlation was statistically significant for GPA.

As per the prediction of student performance using hierarchical multiple regression, results show that Casper has a statistically significant positive effect on two OSCEs, and in many CBFE competencies. This suggests that as Casper scores rise, the scores on those courses are likely to rise as well. There was just one parameter estimate for GPA that was statistically significant ($p < 0.05$);
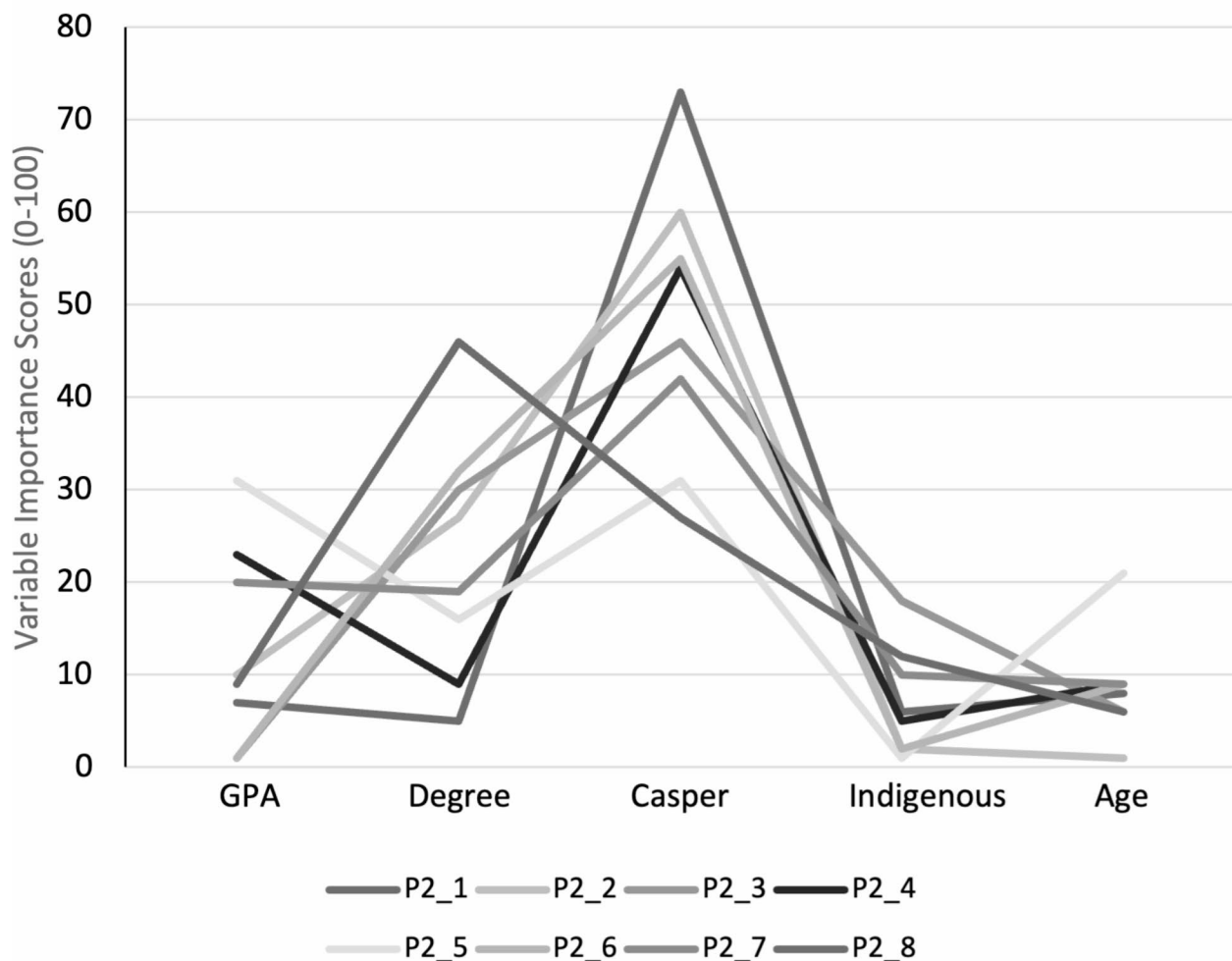
Fig. 4 Variable importance of each admission metric for the competency scores in the second fieldwork placement

however, the parameter was negative, suggesting that the higher the GPA, the lower the competency score on that placement. Also, results showed that adding Casper into the analyses would increase the variance measured in this study.

When looking at the weights of admission metrics identified by non-linear optimization and regression tree analysis, results suggested that the optimal weights of Casper identified by non-linear optimization are much higher than other admission metrics, and GPA is less associated with students' competency scores.

Validity evidence of relationships with other variables - predicting OSCE and fieldwork performances

Our initial assumption was that Casper would predict performance across all OSCEs, because each OSCE is a simulated scenario which requires overt demonstration of professionalism and communication, as well as other aspects measured by Casper such as empathy,

collaboration, and problem-solving. The results partially supported this assumption. The first OSCE is a simulated clinical interview where the focus is on professionalism, verbal communication and demonstration of interpersonal skills, including responsiveness to the client, empathy, and client-centeredness. As expected, Casper predicted performance in the first OSCE over and above GPA. The second and third OSCEs also assess communication and professionalism, but additional emphasis is placed on clinical reasoning—procedure, theory, models, frames of reference, and knowledge of intervention strategies. The added dimensions assessed, in comparison to the first OSCE, place greater relative weight on assessing procedural clinical reasoning, which is concerned with the clinical and evidence-based rationale for the proposed intervention strategy, and less on the demonstration of interactive clinical reasoning in the context of simulated report-back scenarios. The fourth OSCE was similar to a combined first and third OSCE and involved
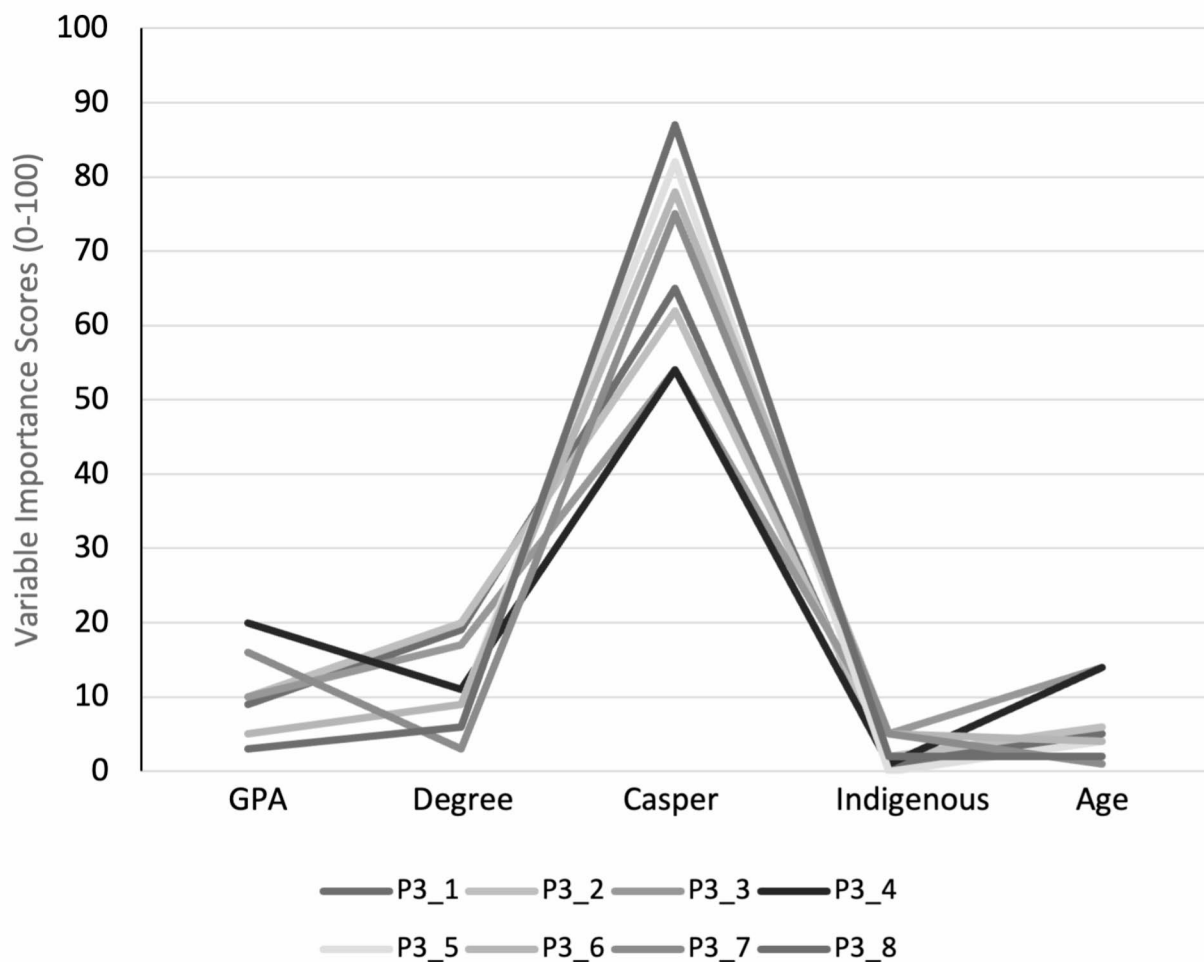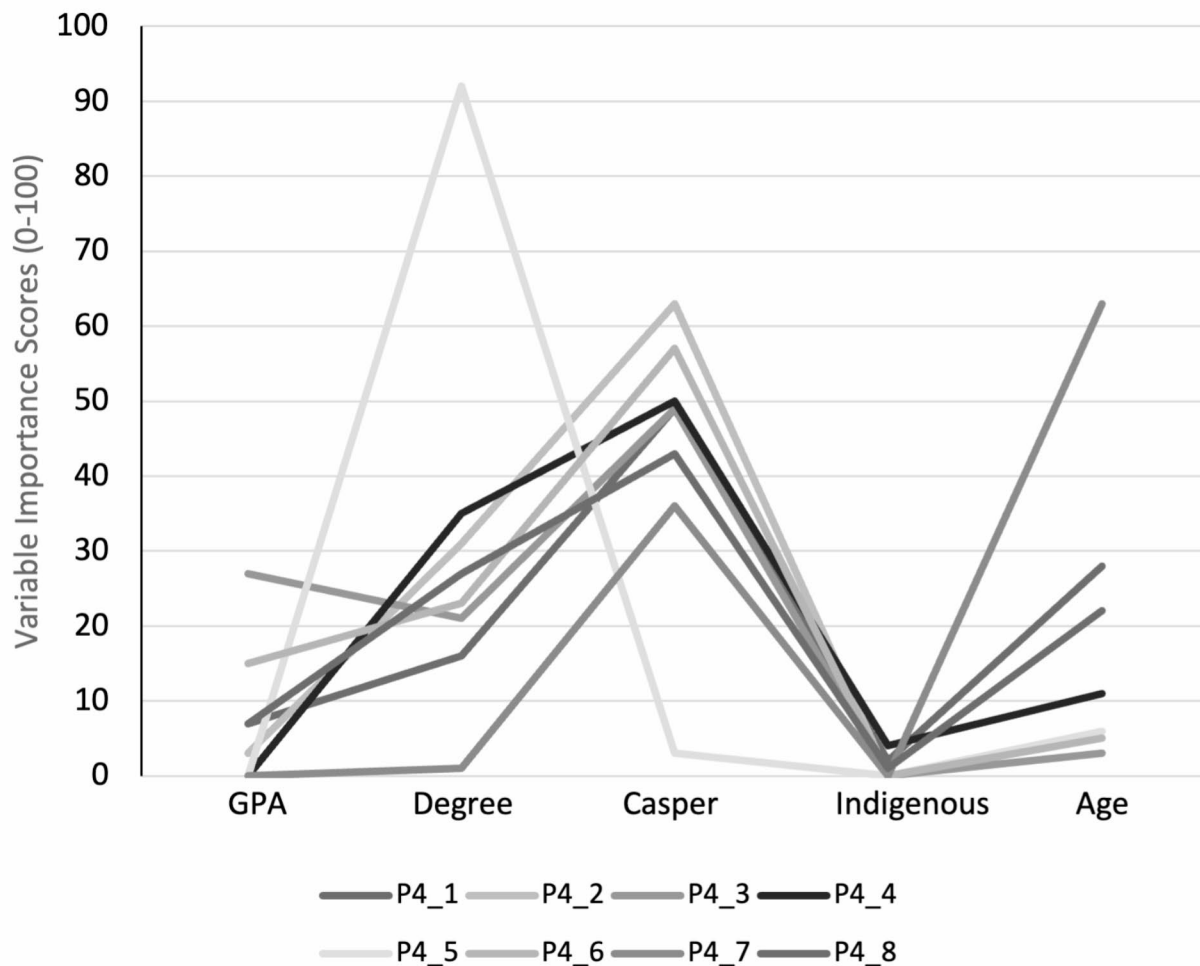
Fig. 5 Variable importance of each admission metric for the competency scores in the third fieldwork placement

interviewing a standardized patient in addition to the communication of assessment findings and proposed intervention strategy. A potential reason that Casper predicted performance on the fourth OSCE is that the interview may have allowed for a more focused assessment of interpersonal skills, which was conducted separately from the communication of assessment results and intervention planning. However, although statistically significant correlations were observed between Casper and the first and fourth OSCEs (i.e., $r_s = 0.19$ and $r_s = 0.26$, respectively), the strength of the associations were relatively weak.

We expected Casper to be a better predictor than GPA of fieldwork performance for all the CBFE competencies, except perhaps for Practice Knowledge which refers to discipline-specific and technical knowledge. The remaining CBFE competencies (i.e., 2–7) all require demonstration of aspects measured by Casper in varying degrees. For example, the CBFE competencies of Clinical Reasoning (CBFE 2), which includes problem-solving

within clinical patient and team encounters, Communication (CBFE 4), and Professional Interactions and Responsibility (CBFE 5), overlap with aspects assessed by Casper such as problem-solving, communication, and professionalism. The results met our overall expectations: Casper was a better predictor than GPA of performance, especially later in the program (i.e., Placements 2 and 3), with statistically significant correlations ranging from $0.19 < r_s < 0.29$. Similar to the observed correlations between Casper and OSCEs, the relationship between Casper and CBFE competencies was weak. However, deviation from this pattern emerged for the final fieldwork placement, where generally, Casper and GPA were poor predictors of fieldwork performance as measured by the CBFE. A potential explanation for this result is the very low variability of CBFE scores due to range restriction. However, results from objective 2, which have fewer data assumptions, converge with those from the regression analysis. In the instance where Casper demonstrated a significant, although weak, negative correlation with

Fig. 6 Variable importance of each admission metric for the competency scores in the fourth fieldwork placement

the CBFE clinical reasoning competency, Casper may better predict what some preceptors interpret the CBFE competencies to represent. This interpretation of the results suggests potential value in examining how preceptors understand and use the CBFE.

Consequences validity evidence - effect of weighting Casper on increasing cohort diversity
Placing greater weight on Casper relative to GPA led to small, non-statistically significant changes in demographic composition between the current group of students and the hypothetical group of selected students, consistent with recent findings in which the inclusion of Casper as an admission metric was not observed to contribute to significant changes in demographic composition [27]. A possible reason for this finding is that the original applicant pool may not be very diverse to begin with. Indeed, advancing a diversity agenda related to admissions will need to start with understanding who is currently applying to our program. This information

could assist the program with setting targets and evaluating the success of strategies, including weighting of Casper in the calculation of an admissions score, for increasing diversity among admitted students. To this end, we showed that the overlap between admitted students, selected primarily on the basis of GPA, and the students who would have been selected using a weighted admission score combining Casper and GPA, was approximately 46%. This result indicates that the intentional inclusion of a measure of personal and professional characteristics combined with GPA may identify a different set of applicants for admission. Notably, the non-overlapping 54% of applicants identified as having both strong academic and non-academic skills, would not have been originally selected and offered admission had the decisions been based on GPA alone. From this perspective, implementation of a weighted admission score to inform selection decisions may better identify individuals with baseline skills from which to build on within the program to become successful health professionals.

Table 7 Weights of admission metrics for each fieldwork placement estimated by regression tree analysis

Competency	Variable Importance	
	GPA	Casper
P1_1	26	**74**
P1_2	22	**78**
P1_3	30	**70**
P1_4	6	**94**
P1_5	32	**68**
P1_6	19	**81**
P1_7	32	**68**
P1_8	6	**94**
P2_1	3	**97**
P2_2	10	**90**
P2_3	38	**62**
P2_4	10	**90**
P2_5	42	**58**
P2_6	10	**90**
P2_7	**51**	49
P2_8	5	**95**
P3_1	3	**97**
P3_2	5	**95**
P3_3	29	**71**
P3_4	25	**75**
P3_5	16	**84**
P3_6	0	**100**
P3_7	17	**83**
P3_8	4	**96**
P4_1	17	**83**
P4_2	16	**84**
P4_3	14	**86**
P4_4	15	**85**
P4_5	13	**87**
P4_6	10	**90**
P4_7	4	**96**
P4_8	27	**73**

Note. The largest weight for each measure of future performance is presented in bold

Table 8 Admission metric differences between admitted students and students selected by weighted admission scores ($N=179$)

Admission metric	Admitted		Weighted Selection	
	Mean (SD)	Range	Mean (SD)	Range
GPA	3.77 (0.12)	3.3–4.0	3.63 (0.21)	2.8–4.0
Casper	0.40 (0.83)	-1.76–2.98	1.10 (0.44)	0.40–2.98
Age	24.09 (2.55)	21–33	24.11 (3.06)	21–50

Our results also suggest that program efforts to increase applicant pool diversity need to be explored further upstream in potential applicants' preparation for the occupational therapy program. It may be an understandable concern for some that, despite the lack of a statistically significant difference, the weighted selection in our

Table 9 Demographic and degree differences between admitted students and students selected by weighted admission scores ($N=179$)

Demographics	Admitted		Weighted Selection	
	N	% of total	N	% of total
Indigenous	7	3.91%	2	1.12%
Female	162	90.5%	163	91.06%
Degree	N	% of total	N	% of total
General arts	29	16%	25	14%
General sciences	48	27%	46	26%
Health	14	8%	29	16%
Kinesiology	61	34%	52	29%
Psychology	27	15%	27	15%

model would have led to fewer Indigenous students being admitted to the program compared with the standard method of admission decision making (2 vs. 7). While caution should be exercised in drawing conclusions about whether this suggests that Casper may be biased against Indigenous applicants, it will be important to monitor whether or not similar outcomes are repeated in future admissions procedures should they incorporate Casper.

Given the history of basing admissions decisions on GPA, in this study we explored whether decreasing the weight of GPA and increasing the weight of Casper would result in lower fieldwork performance. The results show that this was not the case and were in line with the findings from the multiple regression, non-linear optimization, and regression tree analyses. Taken together, the results (see Table 8) suggest that students with a slightly lower GPA (e.g., 3.5–3.6) could be successful in the program as assessed by fieldwork, which is considered to be one of the strongest indicators of success as an entry-level clinician.

Validation of score inferences

This study was positioned within a validity argument framework and evaluated evidence within the extrapolation and implications inferences to support the following central claim of proposed use: *As a measure of personal and professional competencies that inform social intelligence and professionalism, Casper can be used to inform admissions decisions to select a diverse cohort of individuals with a high potential to succeed in practice-based activities requiring non-academic skills.* The results provide preliminary support for the proposed use of Casper with evidence of its relationship with future performance on OSCEs and fieldwork within our occupational therapy program. Our findings align with previous work by Dore and colleagues [2], where Casper predicted future performance in medical education. From a program perspective, and considering the current admissions process, the study results suggest that the inclusion of Casper is value-added as it signals the importance of personal

characteristics, skills, and behaviors associated with social intelligence and professionalism, enhancing the prediction of student performance beyond GPA alone. Notably, the incremental validity reported here (3.2%) was slightly lower than the 5–10% range that was typical of studies in Webster's and colleagues' meta-analysis [18]; however, the studies included in that analysis were limited to training for medicine. An additional value-add to Casper is that it provides an objective measure of personal characteristics, which are important to become an effective health professional, to supplement reference letters and statements of intent. The results also provide preliminary evidence supporting the implementation of Casper to inform admissions decisions, which could lead to the selection of a demographically different cohort compared to selection based only on GPA.

The results of this study point to gaps within the validity argument, namely an evaluation of assumptions related to the scoring inference and how OSCE raters and fieldwork preceptors understand and apply the scoring rubrics and CBFE, respectively. Additionally, re-evaluating the degree to which the domains assessed within the OSCE and fieldwork, as measured by the CBFE, align with Casper may also provide insight into the reasons why the results partially met our expectations.

Limitations of the study

There are several limitations to this study. First, the results are based on one cohort within the context of one occupational therapy program's curriculum and approaches to assessment. For this reason, we advise caution in generalizing our results to occupational therapy programs located at other institutions. Additionally, although we used multiple analytic approaches to cross-validate results, future studies should replicate the analyses with another cohort or aggregation of data across years. Second is regarding preceptor scoring of the CBFE. Especially on Placement 1, some fieldwork preceptors awarded CBFE competency score ratings beyond the scoring range that would typically be expected for students on a first placement. This finding may indicate that preceptors perceived students' capabilities as being much better than the minimum required to master the first placement. Another possibility is that not all preceptors fully understood how to use the rubrics and the target scores for each placement stage. To score students effectively on the CBFE competencies, preceptors need to know how these areas of competency are conceptualized and taught in the curriculum, and to apply that knowledge to render judgements of performance that are situated at a particular time in the students' learning trajectories. Such proficiency provides the foundation for preceptor training by leveraging that knowledge to set appropriate expectations for student performance

and recognize different levels of performance. Further, a qualitative investigation of preceptors' comprehension and use of the CBFE tool with the potential for enhanced training is warranted.

Third, the score ranges recommended within the CBFE for each developmental stage are somewhat restricted and may be considered a limitation of the tool, thus contributing to the low observed correlations. The data for this study entailed scores from performance-based assessments under realistic conditions, which may have also contributed to overall model performance and speaks to the inherent challenge of predicting people's lives. Fourth, while applying a correction for family-wise error (FWE) would be ideal for controlling Type I errors in our analyses, our study, with its relatively small sample size, faced constraints that influenced this decision. Given the limited sample size, using a correction like Bonferroni could have significantly increased the risk of Type II errors, potentially leading to missed detection of true effects. Therefore, to maintain statistical power and identify potential trends, we opted not to apply such corrections. We recognize that this decision may increase the risk of Type I errors. Future research with larger sample sizes should apply more stringent FWE corrections to validate these preliminary findings and mitigate this limitation.

However, the current study makes two scholarly contributions to the health professional education literature. Practically, this study contributes to evidence-based occupational therapy education by demonstrating preliminary validity evidence of Casper to support its use in the admissions process. Study results suggest that the inclusion of Casper during the selection process may add value to the program as it enhances the prediction of student performance in practice-based activities beyond GPA alone. Methodologically, this study demonstrates the use of a relatively novel application of regression tree and non-linear optimization analyses to cross-validate results from multiple regression analysis, thus strengthening the conclusions drawn regarding the relative importance of Casper over GPA for predicting future performance in practice-based activities requiring non-academic competencies. The benefit of regression tree analysis was that it gave us weights for each input that we could use to build a hypothetical admissions model that would maximize students' performance on fieldwork competencies, something that hierarchical regression analyses cannot accomplish.

Conclusion

Assessment of both academic capability and non-academic competencies is important to gain a holistic view of applicants to health professions programs. Casper, as an objective measure of an applicant's non-academic

characteristics, is being increasingly adopted as part of admissions to health professions programs, including occupational therapy, within North America. The results of this study provide preliminary evidence to support the use of Casper as part of the admissions process to an occupational therapy program by examining its associations with future performance in OSCEs and fieldwork over and above GPA. Although the inclusion of Casper improved the prediction of OSCEs and fieldwork, the overall prediction of Casper and GPA is low. Prediction of students' performance in OSCEs and fieldwork is of interest because both OSCEs and fieldwork simulate and/or provide direct experiences requiring the use of non-academic competencies that students are expected to encounter in future clinical practice. Finally, the present study suggests that further research is needed to better understand whether integrating Casper into program admissions may be beneficial for fostering diversity within the student body.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-024-06071-0.

Supplementary Material 1

Data availability
The datasets generated and/or analysed during the current study are not publicly available due to institutional ownership of the data, privacy considerations, and ongoing data analysis but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Ethics approval for this study was granted by the University of Alberta's Research Ethics Board 2. The study's Ethics ID number is Pro00095938. As the study is retrospective in nature the need for informed consent was waived by University of Alberta's Research Ethics Board 2. The research protocol we described as part of our ethics application entailed pseudonymization techniques to anonymize participant data, with password protection to ensure the confidentiality and integrity of personally identifiable information (PII). This secure repository meticulously safeguarded critical identifiers such as Student ID, Class ID, and name and last name, thereby adhering to data protection regulations and mitigating the risk of unauthorized access or data breaches. This study did not involve collecting any new data from human participants. Data used in the study pertained to students' program application packages and pre-existing data gathered during program application or class performance assessments, but was deidentified when it was obtained.

Consent for publication
Not applicable.

References
1. Acuity Insights. Casper technical manual. 2022. https://acuityinsights.com/Casper-technical-manual/
2. Dore KL, Reiter HI, Kreuger S, Norman GR. Casper, an online pre-interview screen for personal/professional characteristics: prediction of national licensure scores. Adv Health Sci Educ Theory Pract. 2017;22(2):327–36. https://doi.org/10.1007/s10459-016-9739-9.
3. Siu E, Reiter HI, Overview. What's worked and what hasn't as a guide towards predictive admissions tool development. Adv Health Sci Educ Theory Pract. 2009;14(5):759–75. https://doi.org/10.1007/s10459-009-9160-8.
4. Grice KO. Use of multiple mini-interviews for occupational therapy admissions. J Allied Health. 2014;43(1):57–61.
5. Shipper ES, Mazer LM, Merrell SB, Lin DT, Lau JN, Melcher ML. Pilot evaluation of the computer-based Assessment for Sampling Personal characteristics test. J Surg Res. 2017;215:211–8. https://doi.org/10.1016/j.jss.2017.03.054.
6. Canadian Association of Occupational Therapists. Canadian Association of Occupational Therapists; 2007. https://caot.in1touch.org/site/pt/codeofethics?nav=sidebar
7. Agho AO, Mosley BW, Williams AM. A national survey of current admission practices in selected allied health educational programs. J Allied Health. 1999;28(1):8–14.
8. Grapczynski CA, Beasley J. Occupational therapy admissions: professionalization and personality. J Allied Health. 2013;42(2):112–9.
9. de Boer T, Van Rijnsoever F. In search of valid non-cognitive student selection criteria. Assess Eval High Educ. 2021;47(5):783–800. https://doi.org/10.1080/02602938.2021.1958142.
10. de Visser M, Fluit C, Cohen-Schotanus J, Laan R. The effects of a non-cognitive versus cognitive admission procedure within cohorts in one medical school. Adv Health Sci Educ. 2018;23(1):187–200. https://doi.org/10.1007/s10459-017-9782-1.
11. Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. Predictive validity of the multiple mini-interview for selecting medical trainees. Med Educ. 2009;43(8):767–75. https://doi.org/10.1111/j.1365-2923.2009.03407.x.
12. Association of Canadian Occupational Therapy University Programs (ACOTUP). Canadian Occupational Therapy University Programs. 2021. https://www.acotup-acpue.ca/English/sites/default/files/Website%20Canadian%20Occupational%20Therapy%20University%20Programs%202021_EN.pdf
13. Juster FR, Baum RC, Zou C, Risucci D, Ly A, Reiter H, et al. Addressing the diversity–validity dilemma using situational judgment tests. Acad Med. 2019;94(8):1197–203. https://doi.org/10.1097/ACM.0000000000002769.
14. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. Evaluations of situational judgement tests to assess non-academic

attributes in selection. Med Educ. 2012;46(9):850–68. https://doi.org/10.1111/j.1365-2923.2012.04336.x.

15. Jerant A, Henderson MC, Griffin E, Rainwater JA, Hall TR, Kelly CJ, Peterson EM, Wofsy D, Franks P. Reliability of multiple mini-interviews and traditional interviews within and between institutions: a study of five California medical schools. BMC Med Educ. 2017;17:1–6. https://doi.org/10.1186/s12909-017-1030-0.

16. Kreiter CD, Yin P, Solow C, Brennan RL. Investigating the reliability of the medical school admissions interview. Adv Health Sci Educ Theory Pract. 2004;9:147–59.

17. De Leng WE, Stegers-Jager KM, Born MP, Themmen AP. Influence of response instructions and response format on applicant perceptions of a situational judgement test for medical school selection. BMC Med Educ. 2018;18(1):1–10. https://doi.org/10.1186/s12909-018-1390-0.

18. Webster ES, Paton LW, Crampton PE, Tiffin PA. Situational judgement test validity for selection: a systematic review and meta-analysis. Med Educ. 2020;54(10):888–902. https://doi.org/10.1111/medu.14201.

19. Lievens F, Sackett PR. The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. J Appl Psychol. 2012;97(2):460–8. https://doi.org/10.1037/a0025741.

20. Adam J, Bore M, McKendree J, Munro D, Powis D. Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. BMC Med Educ. 2012;12(1):69. https://doi.org/10.1186/1472-6920-12-69.

21. Lia C, Cavaggioni G. [The selection process in the Faculty of Medicine: the usefulness of a psychological and aptitude assessment]. Clin Ter. 2013;164(1):39–42. https://doi.org/10.7417/T.2013.1510.

22. Dore KL, Reiter HI, Eva KW, Krueger S, Scriven E, Siu E, et al. Extending the interview to all medical school candidates—computer-based multiple sample evaluation of noncognitive skills (CMSENS). Acad Med. 2009;84(10):S9–12. https://doi.org/10.1097/ACM.0b013e3181b3705a.

23. Lievens F, Patterson F, Corstjens J, Martin S, Nicholson S. Widening access in selection using situational judgement tests: evidence from the UKCAT. Med Educ. 2016;50(6):624–36. https://doi.org/10.1111/medu.13060.

24. Kirch DG. Transforming admissions: the gateway to medicine. JAMA. 2012;308(21):2250–1. https://doi.org/10.1001/jama.2012.74126.

25. Buchs SR, McDaniel MJ. New advances in physician assistant admissions: the history of situational judgement tests and the development of Casper. J Physician Assist Educ. 2021;32(2):87–9. https://doi.org/10.1097/JPA.0000000000000350.

26. Gustafson CE, Johnson CJ, Beck Dallaghan GL, Knight OR, Malloy KM, Nichols KR, Rahangdale L. Evaluating situational judgment test use and diversity in admissions at a southern US medical school. PLoS ONE. 2023;18(2):e0280205. https://doi.org/10.1371/journal.pone.0280205.

27. Leduc JM, Kpadé V, Bizimungu S, Bourget M, Gauthier I, Bourdy C, et al. Black students applying and admitted to medicine in the province of Quebec, Canada: what do we know so far? Can Med Educ J. 2021;12(6):78–81. https://doi.org/10.36834/cmej.72017.

28. Kane MT. Validating the interpretations and uses of test scores. J Educ Meas. 2013;50(1):1–73. https://doi.org/10.1111/jedm.12000.

29. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ. 2015;49(6):560–75. https://doi.org/10.1111/medu.12678.

30. Messick S. Meaning and values in test validation: the science and ethics of assessment. Educational Researcher. 1989;18(2):5–11.

31. Bossers A, Miller PA, Polatajko H, Hartley M. Competency based fieldwork evaluation for occupational therapists. Thomas Limited; 2007.

32. Avriel M. Nonlinear programming: analysis and methods. Courier Corporation; 2003.

33. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Routledge; 2017.

34. Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. Korean J Anesthesiol. 2017;70(2):144–56. https://doi.org/10.4097/kjae.2017.70.2.144.

35. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.

Publisher's note