



**QUEEN'S
UNIVERSITY
BELFAST**

Evaluating selection methods for admission into the social work degree: The Social Work Match psychological test and Prerecorded Virtual Interview. Social Work Match Psychometric

Roulston, A., McNeill, C., Hayes, D., McFadden, P., & Sheilds, A. (2024). *Evaluating selection methods for admission into the social work degree: The Social Work Match psychological test and Prerecorded Virtual Interview. Social Work Match Psychometric.*

Document Version:

Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright 2024 the Authors.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>



**QUEEN'S
UNIVERSITY
BELFAST**



**Identity
Exploration**

**Evaluating selection methods for admission into the social work degree:
The Social Work Match psychological test and Prerecorded Virtual Interview.**

Authors

Professor Audrey Roulston, Queen's University Belfast

Colin McNeill, Identity Exploration Limited

Professor Davy Hayes, Queen's University Belfast

Dr Paula McFadden, Ulster University

Dr Amanda Shields, Open University

January 2024



CONTENTS

Funding and Acknowledgements	page 3
Social Work Match Research Steering Group members	page 3
List of Tables	page 4
List of Figures	page 5
Abstract	page 6
Introduction	page 8
Methods	page 9
Research questions	page 10
Data collection	page 11
Findings	page 16
Validity	page 27
Reliability	page 33
Fairness	page 39
Feedback from participants (Qualitrics survey)	page 41
Conclusion and recommendations	page 46
References	page 48

Funding

This study was funded by the Department of Health.

Acknowledgements

The authors would like to take this opportunity to thank the applicants who participated in the SWM23 study and provided feedback on their experience. They would also like to thank the academic and administrative staff who facilitated recruitment or provided data for analysis, Dr Jana Ross for analysis of Qualtrics data, and members of the research steering group who supported this project.

Research Steering Group members

- Professor Audrey Roulston (Chair), School of Social Sciences, Education and Social Work (QUB)
- Professor David Hayes, School of Social Sciences, Education and Social Work (QUB)
- Professor Marian Traynor, Associate Dean of Education, School of Nursing and Midwifery (QUB)
- Michael Burns, Policy Lead for Learning and Improvement, Social Work and Social Care Workforce, Office of Social Services, Department of Health
- Allen Erskine, Director, Identity Exploration Limited
- Colin McNeill, Director and Research Psychologist, Identity Exploration Limited
- Joanne Lytle, Professional Officer, Northern Ireland Degree in Social Work Partnership
- Tricia Devlin, Professional Advisor, Northern Ireland Social Care Council
- Dr Shirley Boyle, Lecturer in Social Work at Open University (NI)
- Dr Amanda Shields, Professional Lead for Social Work, Open University (NI)
- Siobhan Wylie, Senior Manager, Bryson Care Practice Learning Centre, Bryson Charitable Group

List of Tables

	Title of Table	Page
1.	The 19 social work values and attributes assessed by SWM23	12
2.	The 5 competencies assessed by PVI with sample questions.	13
3.	Scoring rubric for assessing each interview question: indicative answers.	14
4.	Global Score Rating	15
5.	Statistician's metrics: report on 2023 SW interviews	15 & 37
6.	Top 10 VBR scores (SWM) with value theme scores	17
7.	Marginal VBR scores (SWM) with value theme scores	18
8.	The top quartile of PVI scores with Q scores	19
9.	Bottom (marginal) quartile of PVI scores with Q scores	20
10.	Descriptive Statistics Overall Scores on SWM and PVI	22
11.	Regression Model Summary	24
12.	SWM23 and PVI measures compared using basic standards for data quality	26
13.	Evidence of the validity of SWM23 scores	28
14.	Evidence of validity of PVI questions and global scores	29
15.	Unrotated Factor Loadings and Communalities	31
16.	Unrotated Factor Loadings and Communalities PVI Assessor 1	32
17.	Unrotated Factor Loadings and Communalities PVI Assessor 2	32
18.	Correlation Matrix SWM	35
19.	Correlation Matrix PVI Assessor 1	35
20.	Correlation Matrix PVI Assessor 2	36
21.	Was it easy to understand the questions on SWK23?	43
22.	Comparing SWM23 experience to other selection processes.	44

List of Figures

Figure	Title of Figure	Page
1	Sample of the nine-point, semantic differential scale with centre zero	12
2	Distribution of scores on both tests is normal but distinctly different.	21
3	Individual value plot of distribution of overall scores.	21
4	A positive linear relationship between scores on SWM23 and PVI	23
5	A normal distribution of residuals	25
6	Scree Plot for SWM Factors	31
7	Scree Plot for PVI Assessor 1 Factors	32
8	Scree Plot for PVI Assessor 2 Factors	33
9	A positive linear relationship between A's and B's scoring of applicants	38

Evaluating selection methods for admission into the social work degree: The Social Work Match psychological test and Prerecorded Virtual Interview.

Abstract

Oversubscribed social work (SW) courses and a workforce review in Northern Ireland prompted a review of admissions, to ensure recruitment of applicants with strong core values. Calls for values-based recruitment underpinned this research. A values-based psychological screening tool (SWM) was designed and successfully tested using student social workers.

Aims: This study evaluates this screening tool modified slightly for delivery to SW applicants, Social Work Match 2023 (SWM23) and contrasts it with the Prerecorded Video Interviews (PVI) introduced during COVID-19. This study evaluates the SWM23 test as a selection method for admission to SW programmes in Northern Ireland by examining the following:

- (a) the quality of score data produced by the cohort,
- (b) its effectiveness in distinguishing between applicants
- (c) it's psychometric characteristics and
- (d) comparing these SWM23 characteristics with those of the Prerecorded Virtual Interview (PVI).

The study will also

- (f) Use Qualtrics survey software to obtain and analyse feedback on SWM23 by applicants to Ulster University, Queen's University and Open University.

Methods

Applicants who applied for the 3-year undergraduate route (UGR) or the 2-year relevant graduate route (RGR) were invited to participate in the study after completing their admissions interview. Written consent was received from 174 participants, 49 of whom (6 male, 43 female) completed SWM23. The study used a repeated measure design in which the consenting group of applicants are assessed for suitability for the NI Degree in social work courses using two different measures, the virtual interview (PVI) and a psychological test of values (SWM23). A survey collected data on applicants' views of the test.

Findings

Analysis of data provides practical, theoretical, statistical, and qualitative reasons for concluding that SWM23 is a valid, reliable, and fair measure with good discriminatory power, standardized administration, and consistent marking. Furthermore, the PVI is a valid, reliable, and fair virtual interview with good discriminatory power, sound evaluation, and effective technical support consistent with the methodology. However, there are potential issues with unconscious bias towards certain applicants, which are not evident in SWM23. Furthermore, although independent assessors attend annual training and standardization, evidence suggests that marks awarded for responses to questions differ according to the perspective of markers, which is subsequently disguised in the overall mark awarded.

Conclusion

Both measures are valid, reliable, fair, and effective. They complement one other but measure different aspects of suitability for social work training. One measures personal values and the other key SW competencies. Using the measures in tandem would increase the amount of reliable information available to inform the selection process and minimise uncertainty about fundamental suitability.

Recommendations

The findings should be presented at relevant regional committees to demonstrate the benefits of implementing SWM23 as a form of values-based recruitment, prior to any decisions regarding implementation.

If there is a decision to implement SWM23 as a mandatory element of the selection process, universities and partnering agencies would need to agree how the additional data on applicants would be used to inform selection, alongside interview scores, academic qualifications, and any other relevant information determining suitability.

Furthermore, formal agreements should be established between the universities and Identity Exploration Limited, as the latter owns Intellectual Property rights for the SWM23 psychometric test.

Introduction

Research over the past 20 years indicates that students are often motivated to study Social Work (SW) based on altruism, political ideals, or a desire to help people by challenging oppression and promoting a more socially just society (Hackett et al., 2003; Ferguson et al., 2000; Tham and Lynch, 2014; McCartan et al., 2022). Other research indicates that life and family experiences, personal needs and beliefs may motivate people towards a career in SW (Wilson and McCrystal, 2007; Stevens et al., 2012). However, it is widely recognised that SW is a demanding and stressful career, and SW recruitment problems are difficult to resolve (Baginsky, 2013; McFadden, 2018; McFadden et al., 2019).

According to Stratton (2000), the most common methods for selecting SW applicants are academic records, application form, references, interviews, and self-selection. Others report that Personal Statements (PS) are used to gather information about applicants to undergraduate and graduate programmes (Clinedinst, 2019; Klieger et al., 2017; Woo et al., 2020). The PS collects information about motivation to study a particular field and assesses written communication skills (Kuncel et al., 2020). When Cree et al. (2018) reviewed approaches used for admission into SW they indicated that there is little evidence that one method of selection is intrinsically better than another, due to issues around fairness, transparency, and diversity. According to Ryan et al. (2006) the goal for SW programmes is to recruit students who will graduate as competent, ethical, and effective practitioners. Research in Ireland found that SW students who were asked to permanently withdraw from professional training normally failed placements due to professional misconduct linked to SW values (Roulston et al., 2021). When Hayes (2018) formally reviewed complaints made about SWs to the Northern Ireland Social Care Council (NISCC), concerns were highlighted about the honesty and integrity of SWs, who allegedly lied to or deliberately withheld information from service users. Some failed to demonstrate respect or showed bias against service users/carers, or discriminated against them based on age, religion, disability, race, or nationality.

In Northern Ireland (NI) the PS embedded within the Universities and Colleges Admissions Service (UCAS) application was not used for selection. Until 2020, a SW specific PS was used to shortlist SW applicants for interview, which consisted of a 600-word statement responding to regionally agreed questions. Most recently, questions explored their motivation to apply for social work degree training, their understanding of the social work role within a chosen area and the values they bring to social work training. However, concerns were raised about the reliability for shortlisting due to possible plagiarism, coaching or bias. It was concluded the PS was not fit for purpose and that marking did not make best use of resources (Cleland et al., 2012; McNeill et al., 2018; Patterson et al., 2018; Traynor et al., 2019; Roulston et al., 2022).

It has been suggested that assessing the personal characteristics of applicants in a more rigorous manner, using values-based recruitment, may help to improve the selection process (Croisdale-Appleby, 2014; Patterson et al., 2018). However, Patterson et al. (2016) reported that the PS, references, and unstructured interviews were inappropriate for values-based recruitment. 'Social Work Match' (SWM) was developed and piloted with SW students enrolled on a degree programme at Queen's University in 2022 (Roulston et al., 2022). It measured SW values and promoted values-based recruitment. However, an updated version SWM23 was developed for applicants to all SW degree programmes across NI, which was piloted alongside a pre-recorded virtual interview (PVI).

Methods

The study uses a repeated measure design in which the suitability of a consenting group of social work applicants is assessed using two different measures, the pre-recorded virtual interview (PVI) and a psychological test of values (SWM23). PVIs were introduced during the COVID-19 pandemic, to replace face-to-face interviews. SWM23 uses a self-report case-study approach that shares characteristics with three overlapping methods, controlled observation, interview, and psychometric test, to assess the private meaning, relative importance, and emotional significance of an applicant's personal values and provide an overall score on suitability for social work.

This study evaluates the SWM23 test as a selection method for admission to SW programmes in Northern Ireland by examining the following:

- (a) the quality of score data produced by the cohort,
- (b) its effectiveness in distinguishing between applicants
- (c) it's psychometric characteristics and
- (d) comparing these SWM23 characteristics with those of the PVI.

The study will also

- (f) Use Qualtrics survey software to obtain and analyse feedback on SWM23 by applicants to Ulster University, Queen's University and Open University.

Research questions:

- (1) Will the SWM23 instrument provide valid and reliable empirical evidence about the suitability for professional training of an applicant in terms of their SW values?
- (2) Do both the SWM23 test and the PVI meet APA psychometric standards as different but complementary measures of suitability for social work practice?

Descriptive and inferential statistical analysis will use Minitab 21.

Research Setting

Professional SW training in NI is taught at Bachelor's degree level in the following:

Queen's University (QUB), Ulster University (UU) and Open University (OU).

Applicants to Belfast Metropolitan College (BMC) and Southwest College (SWC) are awarded their SW degree through UU. The Department of Health (DoH) currently commissions 275 places per annum across academic institutions. For this study, we recruited a convenience sample of consenting participants who had applied to study at any of the above-named universities.

Sampling and recruitment

On receipt of formal ethical approval from the Research Ethics Committees in each participating university (Ref: 006_2021) and approval from the Regional Admissions Committee, it was agreed that an email would be sent to applicants who completed the PVI. The email included a research project specific URL containing copies of the

approved study documentation (i.e., invitation letter, Participation Information Sheet, and consent form) and a welcome video from the Principal Investigator. A data controller, who was not involved in assessing the PVIs for 2023 entry, obtained written consent from each participant, and sent an email containing a unique identifier and log in details, which allowed access to SWM23 and a short evaluation form. Incentives to participate were also offered (i.e., free entry into a draw to win a £100 and two £50 retail vouchers).

Data collection

With consent from participants, the following information was collected:

- (a) Demographic information (i.e., age, sex, degree programme, postcode).
- (b) PVI question scores (independently marked by two assessors).
- (c) Recorded virtual Interview grades (Global Score 1 - 5) by two assessors.
- (d) SWM23 psychometric scores.
- (e) Outcome of selection process (i.e. offered a place, not offered a place)
- (f) Responses on the evaluation form, captured using Qualitrics.

The Measures

The SWM psychometric

To develop the original version of SWM, Identity Exploration Limited (IEL) engaged with key stakeholders involved in SW admissions across NI, which included regular meetings with representatives from the DoH, the Northern Ireland Social Care Council (NISCC), the NI Degree in Social Work Partnership (NIDSWP), SW agencies in the statutory and voluntary sector, as well as staff, academics, and students involved in admissions to professional degree programmes. Published literature, professional SW standards, and transcripts from qualitative interviews with SW students, academics, and practitioners, were analysed to derive a core set of professional values. These were agreed in consultation with a research steering group. On completion of the original pilot study (Roulston et al., 2022) minor revisions were made to the SWM, in consultation with the research steering group. This improved the suitability of SWM23 for applicants to SW degree programmes.

Table 1: The 19 social work values and attributes assessed by SWM23

Is supportive	treats people with equity	owes respect	is honest	will promote social justice
Is principled	will negotiate	Is accountable	thinks clearly	is a professional carer
Is a listener	Is empathetic	Is empowering	Is realistic	Is committed to CPD
thinks resources are no excuse	will care for self	uses power wisely	Is resilient	-

SWM emerged from the design stage with a set of 19 values preferred by the profession (see Table 1) which come together as a value base (VB) and are reworked into five themes; ‘Professional standards’ (PS), ‘Relationship with service user’ (RSU), ‘Character’ (CH), ‘Resilience’ (RES) and ‘Self-care’ (SC). Each attribute or value is presented as a ‘dimension’ (e.g., is supportive ... is controlling) connecting two contrasting points of view (a construct), presented as a discourse. Respondents use a nine-point, semantic differential scale with centre zero.

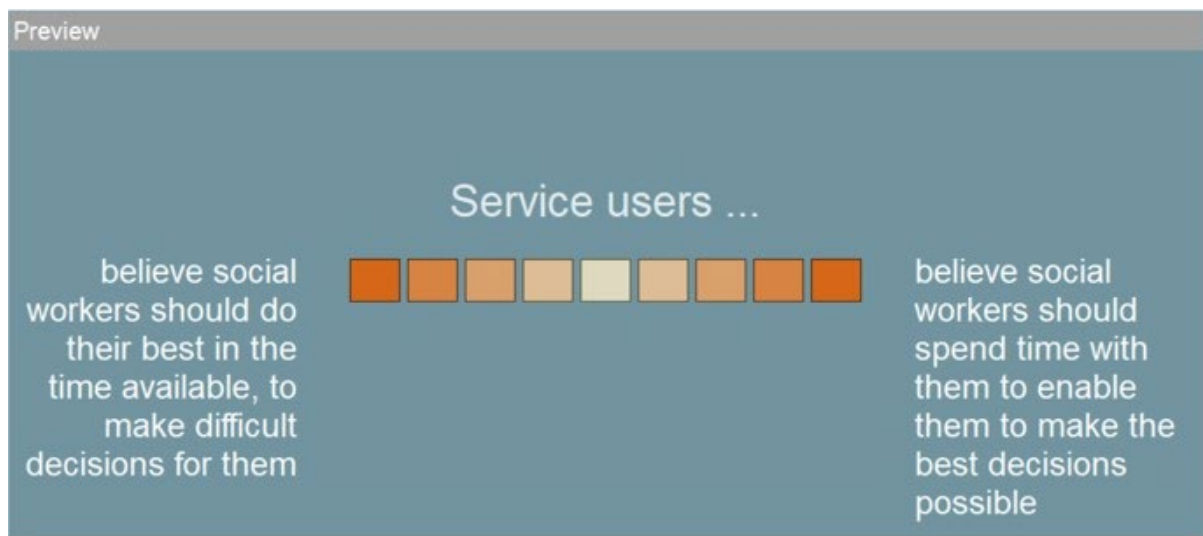


Figure 1 Sample of the nine-point, semantic differential scale with centre zero

One pole of each construct consists of a preferred professional attribute or value. Preferred polar values are presented in a randomized manner. Respondents ‘tag’ the attribute they prefer when they appraise ‘ideal’ or ‘aspirational’ self. Their personal preference may or may not be a professional preference. Respondents are invited to use the centre point scoring zero if they cannot decide between polar values or understand the question.

Ipeus software records the responses from applicants as they use the value constructs to appraise self and entities from the social work domain and reports the outcome as scores on theoretical concepts from ISA, three of which are used to calculate a suitability score for each professional value, an all-value based overall score (VBR) and a score on each theme (composed of relevant sets of values). The cohort of applicants are rank ordered by their overall value based (VBR) score.

The Pre-recorded Virtual Interview (PVI):

Modern Hire provides an online platform to enable applicants to complete a PVI in response to regionally agreed questions, some of which involve the use of case scenarios. The five competencies used to assess candidates applying for the SW degree were chosen, articulated as open questions addressing the competencies and aligned with scoring guides. Broadly speaking, a suitable SW applicant presents as well motivated, with a natural feel for social work and professional values, as well as the ability to self-care and understand why social exclusion or oppression occurs (see Table 2).

Table 2: The 5 competencies assessed by PVI with sample questions.

Question	Competency	Examples of the type of question asked.
1	Motivation	What motivated you to apply for the BSW degree?
2	Case scenario	How would you respond to a nominated case scenario?
3	Social Work values	How would SW values influence your response to the scenario?
4	Self-care	SW training can be stressful. Based on personal experience how will you manage self-care?
5	Diversity and oppression	A situation in which diversity was a factor and social exclusion occurred was outlined. Why do you think it happened?

Candidates are invited to complete the PVI on the Modern Hire platform within an agreed timescale, using their personal computer and video camera. The interview process and instructions are explained beforehand, and they can practice using the

platform. At the start of the PVI, applicants must provide proof of identity. After indicating readiness to start the PVI, questions are presented to applicants on the screen, and they have one minute to read each question and two minutes to provide their answer to each question, all of which are video recorded. In relation to the case scenario, they have five minutes thinking time, prior to answering the question.

Two assessors (one academic and one agency representative), who have completed annual training, are allocated to independently mark a batch of interviews. All assessors use an online assessment form to score each applicant on their response to the five questions (Table 3) and provide a Global Score to represent their overall impression of the applicant’s responses (Table 4). Interview questions are marked out of 50 (maximum of 10 per question). Applicants meeting the threshold of 20 (40%) or above are deemed ‘suitable’ to study social work but would not be guaranteed a place on the course, as the cut-off score for offering places on the UGR and RGR programmes varies each year across the academic institutions. Applicants scoring below the 40% threshold are informed by the Admissions and Access Team that they have been unsuccessful and can request written feedback from the interview panel.

Table 3: Scoring rubric for assessing each interview question: indicative answers.

Classification	Absent	Poor-Limited:	Acceptable - Good:	Very Good	Excellent
Score	0	1-3	4-6	7-8	9-10
Performance	Applicant was unable to answer the question.	Applicant demonstrates poor or limited understanding of or explanation about the competency	Applicant demonstrates an acceptable to good understanding of or explanation about the competency.	Applicant evidences a very good level of understanding and insight about the competency	Applicant shows excellent analytical skills and insight into challenges they may face

Table 4: Global Score Rating

1	Failing, below standard
2	Borderline, not sure
3	Satisfactory performance
4	Good performance
5	Excellent performance

Table 5: Statistician’s metrics: report on 2023 SW interviews

Batch	N	Interviews	Cut score	R-square	Alpha	Exam Var
1	481	962	41.63	0.798	0.884	20.7%
2	147	294	41.66	0.852	0.908	13.9%
3	37	74	39.26	0.849	0.922	24.0%
Total	658	1330			0.891	
With 2 Interviews	7			SD	17.55	
				SEM	5.78	

On completion of the marking, PVI questions scores are forwarded to an independent statistician who assesses the reliability and validity of the question scores and generates a rank ordered list of applicants. Regression analysis is used by the statistician to calculate a cut-off score for the cohort, which represents the anticipated threshold for the admissions staff who make offers. Once the cut-score for each batch is established all candidates are rank ordered, this information is sent to staff in each university’s admissions team. The statistician also notes any aspects of concern about assessors or data for consideration.

Findings

Of the total number of applicants who met the inclusion criteria (N = 658), and were invited to participate, 174 provided written consent (26.4% response rate). However, of those, only 49 (6 male; 43 female) completed the SWM23 data collection (28.16% response rate). Three factors may have contributed to the low uptake in the SWM23 study. Firstly, applicants were not allowed to be invited to participate in the study until they had completed their PVI. Secondly, due to pressure on administrative staff, the invitation to participate in the study was emailed out after first round offers to study SW at UU or QUB had been made, meaning unsuccessful applicants may have been less inclined to participate. Finally, due to the need to allocate unique identifiers and generate login details after each person consented, there was a time delay between applicants consenting and being provided with access to the SWM23 instrument and evaluation form. Despite limitations imposed by the small number of applicants who completed the test (n = 49) the primary research aims were achieved. What follows is the score data for both SWM23 and the PVI with basic and inferred characteristics described in some detail.

Score outcomes for applicant group using SWM23 and PVI

The data sets that follow have been extracted from the complete data sets for those who completed both the SWM23 test and the PVI interview to illustrate the data that admissions staff would receive to inform the selection process.

Table 6 sets out the top ten test scores by rank. The overall ratings (OR) are excellent (5) or good (4). It is not surprising then to see a good standard maintained across the value themes although there are hints strengths and aspects for improvement.

Table 6: Top 10 VBR scores (SWM) with value theme scores

PS	RSU	CH	RES	SC	VBR	Rank	OR*	SUIR	
84.29	83.50	83.24	82.06	82.67	83.42	01	5	10	Value Theme
77.00	78.30	84.80	90.56	82.50	80.67	02	5	37	PS Professional Standards
78.03	79.22	78.29	82.37	77.80	79.86	03	5	19	RSU Responsiveness to Service User
84.88	77.60	73.05	84.96	68.19	78.08	04	5	04	CH Character
79.40	78.00	78.44	78.00	75.63	77.96	05	5	25	RES Resilience
78.99	70.53	80.52	67.81	67.05	72.79	06	5	14	SC Self-care
74.63	72.52	69.79	74.59	68.65	71.45	07	5	30	VBR Value Based (Overall Score)
69.42	69.35	74.08	71.89	65.39	69.83	08	4	29	SUIR Research SUI
69.89	71.16	71.61	67.55	65.98	68.72	09	4	2	
74.25	62.38	67.61	65.29	61.59	67.01	10	4	33	

* OR means overall rating (1 – 5) based on the SD of a score from the cohort mean e.g., 1 is > 1 SD below the mean, 5 is > 1SD above the mean.

By way of explanation, SUIR37 refers to the unique ID code for the participant who scores well on resilience (RES 90.56) but not as strongly on professional standards (PS 77.0). Whereas participant SUIR33 scores quite well on professional standards (PS 74.25) but lower on self-care (SC 61.58) or relations with service users (RSU 62.38).

When two assessors offer contrasting Global Score ratings in response to the PVI, or the overall ratings generate doubts regarding the suitability of some applicants, the SWM ratings offer additional information for admissions, enabling them to select the most appropriate applicant (Table 7).

Table 7: Marginal VBR scores (SWM) with value theme scores

PS	RSU	CH	RES	SC	VBR	OR	SUIR
61.16	60.21	60.37	55.39	50.46	56.63	3	21
57.99	52.14	51.70	63.50	48.30	55.76	3	38
53.61	52.45	54.11	55.41	57.81	54.72*	3	40
57.24	56.12	45.71	59.82	55.04	53.14	3	26
49.00	56.35	56.87	64.05	50.73	52.59	3	39
53.08	49.15	55.45	58.48	45.32	52.45	3	3
49.54	55.39	52.48	58.28	50.08	52.27	3	32
54.60	48.95	39.97	51.08	46.01	48.17	3	36
47.94	46.59	50.50	47.31	43.41	46.11	3	16
48.42	45.03	51.30	51.81	43.92	46.01	3	31

* Mean 50.54: Median 54.72

As illustrated in Table 7, all applicants have an overall rating (OR) of 'satisfactory' (3) but the VBR ranking distinguishes them. This is explained by their different scores on professional standards, relationships with service users, character, resilience, and self-care.

In relation to the PVI scores, Table 8 outlines the double level of scores awarded by Assessor 1 (upper line) and Assessor 2 (lower/shaded line), based on their assessment of applicant responses to each PVI question. The table also demonstrates the Global Scores awarded by each assessor, which illustrate variation with regards to perceived performance. Differences between assessors can be seen in scores awarded for interview questions and Global Scores. For example, SUIR 38 has an overall PVI score of 79, and is ranked 9th / 42 but the global score with one marker is 5 (excellent performance) and with the other is 3 (satisfactory performance). As the Global Score is not currently used by the university admissions staff, to inform selection, it's value or purpose is unclear. Since the removal of the Personal Statement scoring from the selection process, admissions staff would benefit from additional sources of data on applicants, particularly when the PVI ranking and PVI overall scores are equal, as illustrated with SUIR 38 and SUIR 04, and they need to decide who to approach next on their waiting list.

Table 8: The top quartile of PVI scores with Q scores

SUIR	PVI Overall score	Q1	Q2	Q3	Q4	Q5	PVI Global score	PVI rank /42
15	88	6	6	8	8	9	5	1
		10	10	9	10	10	5	
08	85	7	9	8	9	7	5	2
		9	8	8	8	9	5	
07	85	8	9	6	9	9	5	2
		7	9	7	9	9	5	
21	85	10	8	9	9	10	5	2
		8	8	6	5	9	5	
27	84	8	8	7	7	7	4	5
		10	10	8	8	8	5	
31	82	9	9	6	7	8	5	6
		9	7	9	7	8	5	
17	81	9	7	9	7	8	4	7
		9	8	6	7	8	4	
25	80	7	6	7	8	8	4	8
		9	8	7	8	9	5	
38	79	8	9	8	7	9	5	9
		9	8	6	6	5	3	
04	79	9	9	7	9	5	5	9
		8	8	5	7	8	4	
top quartile Q3: 76.75								

Table 9: Bottom (marginal) quartile of PVI scores with Q scores

SUIR	PVI Overall score	Q1	Q2	Q3	Q4	Q5	PVI Global score	PVI rank /42
41	65	8	5	8	6	5	4	29
		6	5	7	5	4	3	
40	62	6	6	6	7	6	3	30
		6	5	5	5	4	3	
28	61	6	7	6	6	7	4	31
		6	5	3	4	5	2	
36	59	7	5	2	3	6	3	32
		8	7	3	5	6	3	
39	59	7	5	5	4	6	4	32
		5	7	3	6	4	3	
35	62	8	5	5	4	5	2	34
		8	7	5	3	6	3	
37	56	10	8	8	0	0	4	37
		7	6	4	3	3	3	

Basic characteristics: Score data on SWM and the PVI.

Is there a difference? Are the two sets of scores significantly different? The answer is yes; see Figures 2 and 3 for a basic visual comparison. These 42 applicants as a group¹ scored significantly better based on the PVI data (M = 68.07, SD 12.09) than they did on the SWM23 test (M = 51.10, SD = 20.43) resulting in a mean decrease per applicant in the score on the test (M= -16.98, SD = 19.03): see Table 10 for the set of descriptive statistics on SWM23 and PVI. This difference was statistically significant, $t(82) = -5.78$, $p = 0.000$ two tailed. So, the central tendency of the score

¹ Excluding the 6 OU applicants whose scores on the PVI were calculated in a different manner and one other with data issues.

data is different, and this is presented visually in Figures 1 and 2 where the near normal distribution of scores on both PVI (AD 0.373, p-value = 0.403) and SWM23 test (AD 0.441, p-value 0.276) can be seen. The mean and median are close and central in SWM23 (M = 51.0, Med = 53.14) and PVI (M = 68.07, Med = 69.0) so that both have a clearly established central tendency some distance apart.

There is a slight negative skew in the test scores (Skew = -0.69) and the interview scores (Skew = -0.65) and slight positive kurtosis in both (Kurtosis = 0.37 and 0.59 respectively). The effect of left skew is slight (-0.5 to + 0.5 is regarded as symmetrical) and we can consider our data to have zero skew for practical purposes. Test scores often follow a left-skewed distribution with most respondents performing relatively well and a few performing far below average (Turney, 2023). The light tail of low scoring applicants is indicated by the modest Kurtosis figures.

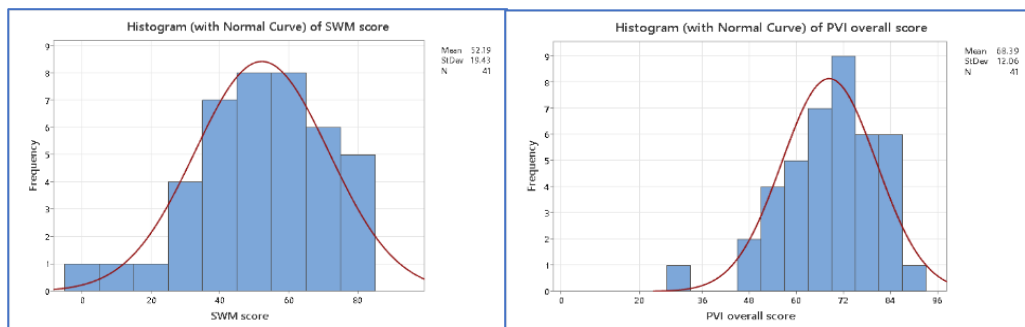


Figure 2: Distribution of scores on both tests is normal but distinctly different.

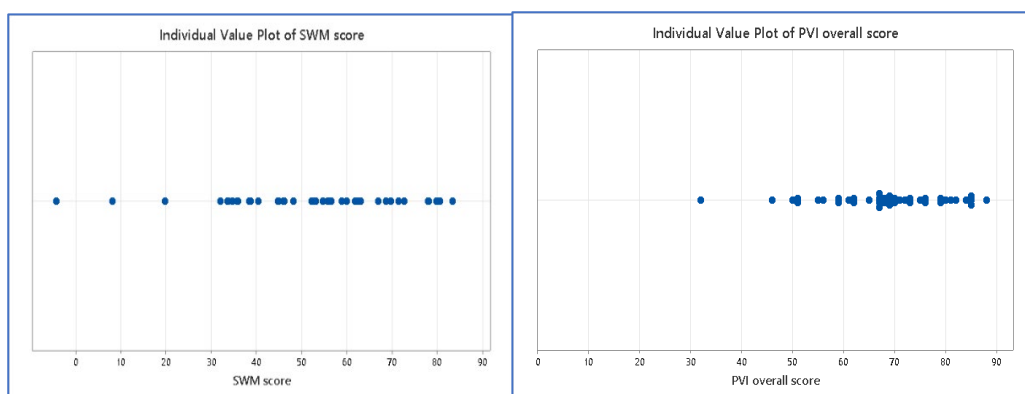


Figure 3: Individual value plot of distribution of overall scores.

Table 10. Descriptive Statistics Overall Scores on SWM and PVI

Variable	SWM (n = 42)	PVI (n = 42)
Mean	51.10	68.07
SD	20.43	12.09
Minimum	-2.16	32.00
Q1	37.78	60.50
Median	52.86	69.00
Q3	66.95	76.75
Maximum	83.42	88.00
Range	85.58	56.00
Mode	*	67 (n = 4)
Skewness	-0.69	-0.65
Kurtosis	0.37	0.59

There were no outliers, all scores fell within the normal distribution on both measures using a Grubb's test, SWM23 ($G = 2.61$, $p = 0.282$), PVI ($G = 2.98$, $p = 0.068$).

A wider range of values is captured by SWM23 (range 85.58) than the PVI (range 56.00) indicating the greater power of SWM23 to discriminate between individuals. The greater variance, degree of variability or dispersion, of the test scores (SD = 20.43) compared with the PVI scores (SD = 12.09) is another indication of the better sensitivity of the test to difference between applicants.

The number of ties in the PVI results, the bunching of same scores, is also evident from looking at Figure 3 emphasised by a cluster of 4 scores at the mode (mode = 67). There is no mode on the SWM23 test scores and there are no ties. As explained elsewhere, the ties that occur using the PVI are managed by the statistician during regression analysis for the benefit of the internal selection process, whereas the results received by applicants show ties.

Inferred characteristics of the score data

Is there a relationship? Using inferential statistics, we confirmed that there is a difference between the results produced by the PVI and the SWM23 test on

suitability for SW but found that there is a moderate positive relationship between them as measures of suitability.

A related t-test was performed on the data from the two measures. The null hypothesis that the measures will produce the same results about suitability for the SW degree and SW practice is rejected and the test is statistically significant: $df = 82$ critical value at $\alpha = 0.05$ is 1.980 ($t = -5.78$) p -value 0.000. A Pearson's correlation was performed to investigate the relationship between scores for suitability using the PVI compared to the SWM23 test. Results of the SWM23 test showed that there was a moderately positive relationship between the two variables ($r = 0.408$, $p = 0.007$). As a rule of thumb about the strength of a relationship based on r , $r = 0.40 - 0.59$ is considered a moderate relationship (Evans, 1996). A regression analysis was performed on the data and established that there was a positive linear relationship (see Figure 4). One measure can, to some extent, predict the scores on the other. The regression equations are: $SWM\ score = 4.15 + 0.6897\ PVI\ overall\ score$, $PVI\ overall\ score = 55.73 + 0.2416\ SWM\ score$. In linear regression a *residual*, (see Figure 4), is the difference between the actual value and the value predicted by the model ($Y - \hat{Y}$) for any given point.

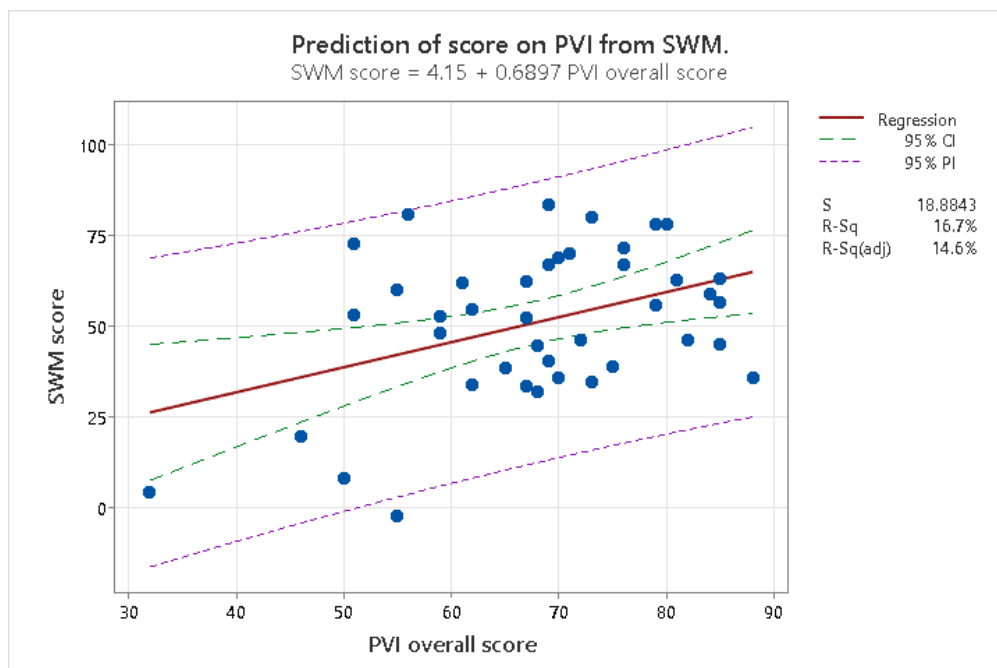


Figure 4: A positive linear relationship between scores on SWM23 and PVI

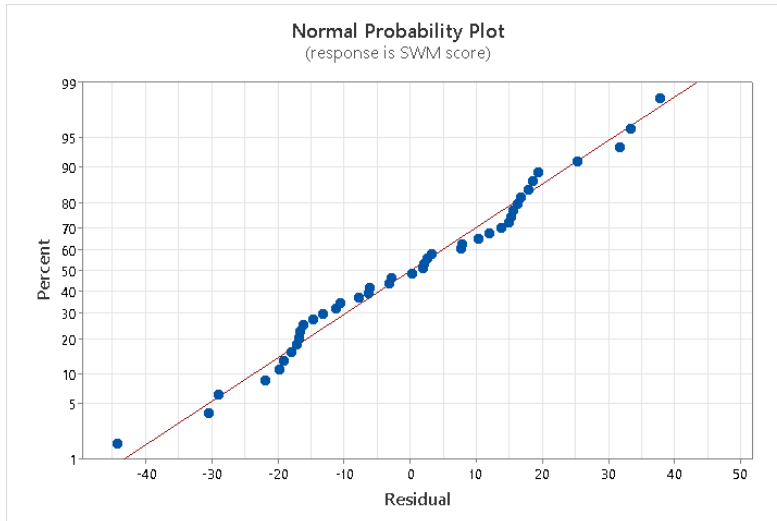


Figure 5: A normal distribution of residuals

Table 11: Regression Model Summary

	S	R-square
SWM23	18.8843	16.66%
PVI	11.1776	16.66%

S = Standard error of the regression line

This is a positive relationship of moderate proportions as reflected in the regression model (Table 11). Conveniently S ($S = 18.8843$) tells us how wrong the regression model is on average using SWM units. S is the standard error of the regression and, as it is also the standard error of the estimate, represents the average distance that the observed values fall from the regression line.

R-squared provides the relative measure of the percentage of the dependent variable variance (16.66%) that the model explains. R-squared tells us how well a linear model fits the data, a 'goodness of fit' statistic.

Summary of findings on data characteristics

Both the SWM23 test and the PVI data have good psychometric characteristics even if they fall short of perfection. The scores have distributions that differ in their norms (mean, standard deviation) but are approximately normal, helping to interpret an individual's score relative to the cohort of applicants.

The data from both tests is internally reliable including (within the PVI) the inter-rater reliability. The tests are both valid on qualitative analysis, although some limitations are indicated with the PVI, and both are shown on statistical analysis to measure essentially one variable or factor called "suitability for the SW degree and SW practice". They measure different overlapping sets of variables within this factor.

In short, SWM23 and PVI are both effective measures of suitability for social work, complying with generally accepted standards for quality. The score data have different distributions and norms for they measure aspects of suitability in different ways, but both distributions are normal. The scores on the measures have a moderate positive linear relationship probably because the SWM23 test only measures SW values and the PVI takes SW values into account.

Quality of score data

As outlined above, the research data was obtained using two distinctive methodologies: SWM23 and the PVI. Although the information sought and provided is different, both methods provided valuable information about suitability for social work (Table 12). The SWM23 assesses meaning and feeling about a comprehensive² set of *SW values*, the PVI assesses five key *competencies* based on experience of the selection process. While APA standards (APA, 2014; 2020) may not be directly applicable to evaluating the PVI, it is still important to ensure their quality. When investigating the quality of data generated by both our measures, three essential standards for psychological tests were applied: (a) validity (b) reliability and (c) fairness.

² Research based

Table 12. SWM23 and PVI measures compared using basic standards for data quality

Criteria for quality	Social Work Match (SWM)	Prerecorded Virtual interview (PVI)
Discriminatory power	<i>Good.</i> Automated decimal place scoring by a software algorithm is used to rank order applicants and rate them 1 – 5 based on SD of their score from the cohort Mean.	<i>Good, conditionally.</i> A subjective <i>global score</i> rates applicants into categories 1 – 5 (whole numbers, integers). A question based <i>overall score</i> (integer) is converted post hoc into a decimal score during regression analysis, which minimises ties in ranking.
Reliability (is it consistent)	Yes. Records in real time the use of nineteen key SW values to appraise multiple entities from home, work, and social domains. Responses are stored immediately, and based on those responses, SW values are scored automatically and so reliably, for stability and affect in an algorithmic process.	Yes, <i>conditionally</i> . An assessor scores a two-minute recorded response to each of five SW competencies and then rates the overall performance in one separate ‘global score’. Two independent assessors appraise applicants in three batches (N = 658 in 2023). Reliability is founded on the quality and consistency of the assessors judgement and is evaluated statistically.
Validity (does it measure what it claims to measure – suitability for BSW degree)	Yes, the software records the use of established professional SW values to appraise self and people or entities known to the applicant by way of responses on a scale indicating the extent to which the value is believed to be true of self, person or entity.	Yes, <i>conditionally</i> . Suitability is assessed by open questions to an applicant addressing five competencies chosen by the NIDSWP, permitting open and honest replies if the applicant is willing to provide them - and not simply use knowledge gained about social work during preparation for interview to promote self-interest in an inappropriate manner.
Standardised (the responses are put on the same scale for comparison purposes)	Yes: work is done in the algorithms to adjust for different styles of use of the SDS scale and standardise outcomes, making individual scores directly comparable. The question of subjective bias in scoring does not arise.	Yes. <i>Conditionally</i> . There is no scale. An impression is left in the assessor’s mind. The assessor ‘meets’ the applicant virtually and may respond differentially to different ‘persons in action’. Any shift in weighting of scores due to unconscious bias, favourable or otherwise, is not easily identified or quantified. There is no exchange of views as in a face-to-face interview that might change this. The assessor can only do their best to be consistently objective in the application of the marking rubrics and the global rating.
Standardised (administration)	Yes: administration and scoring are systematic and consistently the same for all.	Yes: <i>conditionally</i> . The same, systematic, and consistent, process of administration and scoring of questions is used by assessors for all applicants. Overall performance (the global score) is rated (scored) subjectively by two independent assessors and both ratings used.

PVI is a distinct form of assessment, which relies on assessors independently watching and scoring each applicant's 10-minute video in which applicants have responded to five interview questions. The assessor also tries to determine each applicant's overall suitability for SW training (failure, borderline, satisfactory, good, or excellent) without meeting or speaking to them.

Initially, if we consider some basic criteria, we can see that both selection methods generate useful information that can be scored (Table 12). However, there are differences in the nature and the quality of that information. Broadly speaking the quality of the data and efficacy of the PVI is conditional, being dependent on the competence and objectivity of the assessors and the openness and communication skills of the applicants. The SWM23 psychological test is clinical and objective since interpretation of the raw data (the applicant's responses) and scoring is done using algorithms.

1a. Validity

Validity refers to the extent to which a test measures what it claims to measure. Both classic concepts of validity and the evidence based 'unitary' concept are used in this analysis (Tables 14 and 15). *Unitary concept of validity*: appraising validity as a unitary concept means considering evidence based on a network of interrelated factors (nomological net): origin Cronbach and Meehl (1955) and Messick (1995).

Table 13: Evidence of the validity of SWM23 scores

Evidence about validity of SW value scores using SWM²³	
CLASSIC VALIDITY CONCEPTS	Test content and construct validity (of values characterising 'Suitability for SW') is based on a literature search, content analysis, codification, frequency of use count, trialing by SWs, advice and guidance from experienced SWs and a steering committee, re-affirmed by feedback on its use from student social workers and statistically via item and factor analysis (unidimensional - 'Suitability for SW'). The focus on core personal values means that the assessment is valid 'bottom-up', the applicants view of self, social work and the social world.
Content validity	
Construct Validity	
Face Validity	
UNITARY CONCEPT	
Credibility	A direct measure of stability and emotional significance of values
Fitness	Values and entities are fit for purpose by design (classic concepts)
Robustness	Procedure and calculations are systematic and inherently robust
Reliability	The test provides a reliable snapshot of values at time of response
Integrity	Easy to complete and understand, genuine response = integrity
Representativeness	Standardised to make individual scores comparable with others
Coherence	Concepts defined by algorithms for coherent and holistic analysis
Transparency	Data is direct from respondent via interface, analysis is automatic

SWM23: The evidence on this analysis (Table 13), is that scores on personal SW values and the overall suitability score (VBR) were valid because credible, fit for purpose, robust, reliable, had integrity, were representative, coherent, and transparent. However, broadly speaking, this method is a 'bottom-up' perspective, the applicant's (self-report) view of self and social work.

PVI: The PVI measure was also found to be credible, fit for purpose, robust, reliable, and coherent, a valid characterisation of suitability for social work (Table 14).

Although there is no face-to-face interaction applicants, interpersonal communication skills and critical thinking can be demonstrated through the applicant’s response to each question, and the assessor’s overall impression using the Global Score.

Conscious or subconscious bias is a potential source of variance which is monitored and assessed post hoc (see Reliability).

Table 14. Evidence of validity of PVI questions and global scores

Evidence about validity of SW question and global scores using PVI	
CLASSIC CONCEPTS	Face and content validity of the PVI competency questions characterising suitability for SW are based on the professional judgement of the experienced professionals who created them. The SW competency questions are open questions written to address the specific, relevant, aspects of suitability indicated in the marking guides for each question and have construct validity. For example, a marking guide might say, “Very good. Evidences a very good level of understanding of the professional roles and responsibilities of a social worker”. Certainly, “understanding SW roles” is a valid attribute theoretically but measuring its achievement empirically, as the assessor must do, is valid only if the assessor is a sound judge of this attribute. We found evidence suggesting sound judgement by the PVI. The assessment is valid top-down, that is from the university’s perspective.
Content validity	
Construct validity	
Face validity	
UNITARY CONCEPT	
Credibility	Five questions give applicants the opportunity to freely express themselves about their key SW competencies and reviewing their overall performance permitted broader indications of suitability to emerge. Time constraints, the artificial (online) context in which questions are posed, answered and the recorded, ‘distant’, format impose limits on understanding an applicant’s suitability that are not there in a face-to-face interview.
Fitness	The online platform and administration for data collection and delivery to the assessors for marking is fit for purpose although reading the questions from a screen and recording a reply to an anonymous assessor while being video recorded under time pressure is not a fully effective substitute for a face-to-face interview. Marking of questions and the global rating was fine.
Robustness	The PVI procedure is robust, well structured, and systematic, underpinned by a high level of internal reliability of marking by both assessors.

Reliability	Inter-rater agreement is good (it is monitored statistically; internal reliability of scores on questions is high, alpha = 0.89 for cohort) Table 5.
Integrity	The integrity of the PVI process is fine if the applicant's responses are genuine. The assessor does not have the opportunity to clarify, test, or challenge suspected lack of integrity which is a limitation of the PVI. In person face-to-face interviews assessors could only prompt applicants using "anything further to add?" Lack of a face-to-face element imposes a limitation to the credibility of the PVI.
Representative	Suitability for SW is effectively represented by 5 personal characteristics, motivation, self-care, and attitude to diversity, natural affinity for social work and insight into values. Natural affinity for SW and insight into personal values are too broad to be explored in a 10-minute interview but the impression gained by the applicant's performance offers a 'first impressions' representation of suitability for a place on the SW course.
Coherence	The criteria employed were chosen from a much larger set of relevant competencies but are coherent and interconnected providing a holistic 'picture' of the applicant enabling a decision about suitability to be made.
Transparency	The open-ended, free response questions facilitate transparency and the disclosure of relevant information so that informed decisions can be made. But see above on <i>integrity</i> .

Presenting only five SW competencies as representative of suitability narrows the scope of the interview and there is no opportunity to explore any one of these in depth given the nature of the interview process. As is indicated in Table 14, this can affect assessment of integrity of response and interacting with any transparency limitations can reduce clarity of judgement and effectiveness of the interview. *Coherence* was good but was coherence of a limited range of competencies.

Factor analysis

A principal component factor analysis of the SWM23 theme score and the PVI question score correlation matrices indicated a single major factor present in both measures with that factor accounting for more of the variance in the SWM23 scores (97%) than the PVI scores (Table 15). That is a single factor which we name as "suitability for the SW degree and SW practice".

Table 15. Unrotated Factor Loadings and Communalities

Theme	F1	F2	F3	Communality
PS	0.984	0.106	0.112	0.992
RSU	0.994	-0.007	0.006	0.988
CH	0.982	0.138	0.106	0.994
RES	0.984	-0.112	0.086	0.988
SC	0.983	-0.125	0.099	0.992
Variance	4.8549	0.0586	0.041	4.9544
%Var	0.971	0.012	0.008	0.991

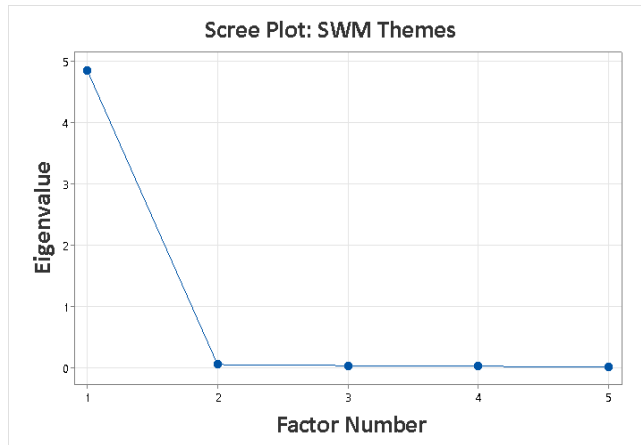


Figure 6: Scree Plot for SWM Factors

A factor analysis was conducted on the scoring of the five PVI questions by assessors, which indicated one principal factor in both cases, but a clear difference in the amount of variance accounted for by that factor (Assessor 1: 57% of variance, See Table 16: Assessor 2: 73.9% of variance, Table 17). Using the established Kaiser guidance on the eigenvalue (amount of variance accounted for) which would justify factor status, that is an eigenvalue of > 1 , we concluded that one major factor was indicated. Factor 2 had an eigenvalue of less than 1, although there is a hint of a second factor being involved in the judgement of assessor 1 (see Figure 7).

Table 16: Unrotated Factor Loadings and Communalities PVI Assessor 1

Theme	F1	F2	F3	Communality
Q1 A	0.786	-0.344	-0.4	0.897
Q2 A	0.829	-0.281	-0.196	0.804
Q3 A	0.676	-0.419	0.568	0.955
Q4 A	0.755	0.492	0.267	0.884
Q5 A	0.72	0.576	-0.15	0.873
Variance	2.8515	0.9459	0.6146	4.412
% Var	0.57	0.189	0.123	0.882

It seems reasonable to suggest that the prime factor, the major source of variance in both measures, be labelled as suitability for social work. This evidence, based on observation and measurement, argues for the validity of both SWM23 and the PVI as measures of suitability for social work.

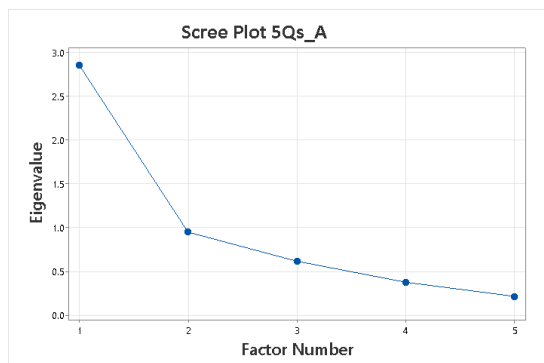


Figure 7: Scree Plot for PVI Assessor 1 Factors

Table 17: Unrotated Factor Loadings and Communalities PVI Assessor 2

Theme	F1	F2	F3	Communality
Q1 B	0.888	0.029	-0.237	0.846
Q2 B	0.862	-0.251	-0.277	0.883
Q3 B	0.81	0.565	0.07	0.98
Q4 B	0.832	-0.211	0.497	0.984
Q5 B	0.902	-0.101	-0.024	0.824
Variance	3.6932	0.4375	0.3853	4.516
% Var	0.739	0.087	0.077	0.903

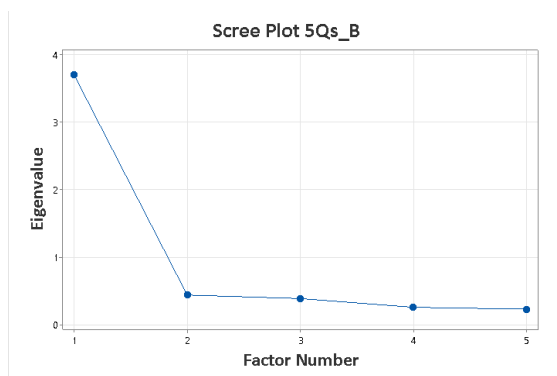


Figure 8: Scree Plot for PVI Assessor 2 Factors

In conclusion, the focus of SWM23 on measurement of nineteen core personal values means that the assessment if valid, is valid ‘bottom-up’. It is the *applicant’s appraisal* of self and SW and their suitability for SW, ingrained in the values they use to make sense of the world. The PVI assessment of five key personal characteristics if valid, is valid top-down, based on what is observed in a ten-minute PVI, independently appraised by two assessors, from different perspectives (the academic and the practitioner), who are well informed about SW training and practice. That is an assessment from the university or *employer perspective*.

While the perspectives on suitability provided by the test and the virtual interview differ, they overlap in content, and both are valid and complementary.

1b. Reliability

An essential notion here is consistency of measurement. Ideally, all applicants should be assessed against the same standardised measure with valid basic units, such as a test scale. However, such consistency is difficult to achieve in a full day of face-to-face interviews (n=10-12 applicants), as assessors may become fatigued as the day progresses or may be influenced by their co-assessor. In relation to PVIs, each recording is viewed privately and independently, and the assessor can pace their marking and take regular breaks to minimise the risk of fatigue. Furthermore, although all assessors are qualified SWs, they come from different perspectives. One assessor will be employed as an academic in a university and will be accustomed to teaching and assessing students over the course of their two- or three-year degree course. The other assessor will be employed by a SW agency (i.e. statutory or voluntary sector, youth/criminal justice, education authority), and may only encounter students when they go on practice placements or graduate, which is

much further into their professional training. Although all assessors complete mandatory training, which includes a short standardization exercise, the different perspectives or expectations of applicants may influence scoring of the responses to questions and the overall Global Score of the applicant. This may help to explain the differences in internal and inter-rater reliability that emerge below. Reliability also implies the production of stable results over time but has not been evaluated in this one-off use of the measures.

Determining the *reliability* of SWM23 and the PVI as selection methods is about consistency of measurement. SWM23 uses a scale, simple direct questions and calculations using algorithms to score the responses. The PVI attempts to be consistent with the provision of mandatory standardisation training and a marking guide for assessors to use when scoring the applicant's response to each question (Table 3). As outlined above, markers score responses to five open questions, which are carefully worded to evoke insightful answers and evidence of broad competencies. Two assessors independently consider the ten-minute PVI and award a Global Score Rating to classify performance (Table 4).

The nomination and training of competent assessors is one key to the reliability of results and the reliability of their judgement is carefully monitored. Post hoc the reliability of the results is analysed and formally tested by a statistician for *internal* and *inter-rater reliability* with feedback to admissions staff (Table 5). The PVI questions and the scoring rubric are reviewed annually, with a different set of PVI questions used for each round of interviews.

One can assess the degree to which a measurement tool is reliable in several ways. The most relevant in this context being *internal consistency* (item or scale reliability) and *interrater reliability*. There is secondary evidence of *test-retest* reliability for the SWM test (Roulston et al., 2022) but no such evidence on which to evaluate test-retest reliability of the PVI.

Internal consistency (item or scale reliability)

SWM23: Internal consistency of the SWM23 test was investigated using a Pearson Product-Moment Correlation Coefficient and Cronbach's alpha. The high, positive

values in the Correlation Matrix (Table 18) indicate that all items are highly correlated with each other. The matrix plots (available upon request) indicate that all the items have a linear and positive relationship. The overall Cronbach's alpha is 0.9925, which is greater than the common benchmark of 0.7. Therefore, we conclude that all the SWM value themes are measuring the same characteristic (suitability for SW).

Table 18: Correlation Matrix SWM

	PS	RSU	CH	RES
RSU	0.975			
CH	0.963	0.972		
RES	0.958	0.976	0.950	
SC	0.951	0.975	0.953	0.965

Cell contents Pearson correlation

PVI: A Pearson Product-Moment Correlation Coefficient and Cronbach's alpha was used to investigate and compare the internal consistency of the scores on answers to the five PVI questions as marked by assessors 1 and 2³. Internal consistency of the PVI changed when the assessor changed.

Assessor one's marking of responses to questions had a relatively high level of internal consistency ($\alpha = 0.8009$) indicating that the question items are interrelated and provide consistent measurements of suitability for social work (Table 19). A Pearson correlation was used to investigate the correlation between scores on individual question items which were all positive but varied from weakly positive ($r = 0.233$) to strongly positive (0.694) see the Matrix in Table 19.

Table 19: Correlation Matrix PVI Assessor 1

	Q1	Q2	Q3	Q4
Q2	0.694			
Q3	0.473	0.503		
Q4	0.327	0.478	0.391	

³ It is assumed that the same two assessors were used for each of the three batches, but this has not been confirmed.

Q5	0.436	0.398	0.233	0.667
----	-------	-------	-------	-------

Cell Contents Pearson correlation

Cronbach's Alpha
0.8009

On assessor two's marking of responses to the five PVI questions the virtual interview had a high level of internal consistency ($\alpha = 0.9111$) indicating that the question items are interrelated and provide consistent measurements of suitability for SW (Table 20). A Pearson correlation was used to investigate the correlation between scores on question items. These were all positive but varied from moderate positive ($r = 0.585$) to strongly positive correlation ($r = 0.758$), see the Matrix in Table 20.

Table 20: Correlation Matrix PVI Assessor 2

	Q1	Q2	Q3	Q4
Q2	0.731			
Q3	0.671	0.585		
Q4	0.642	0.644	0.590	
Q5	0.758	0.737	0.654	0.709

Cell Contents Pearson correlation

Cronbach's Alpha
0.9111

An independent statistician, employed by QUB and UU provided metrics on the results for the cohort of 2023 applicants (Table 5) ($N = 658$) which encompasses all the PVIs broken down into three batches of interviews. A central aim of the metrics is to test the reliability of the assessment process and provide feedback if concerns arise. Item reliability (internal consistency) for each batch of applicants interviewed, including results from both assessors was good or excellent ($\alpha = 0.884, 0.908, 0.922$) and for the entire cohort of 1330 PVIs was good ($\alpha = 0.891, > 0.90$ being excellent). These coefficients are of the same order of magnitude we found with our

small sample ($\alpha = 0.8009$ and 0.9111) and complement the above findings for reliability.

Statistician’s metrics: report on 2023 SW interviews (Table 5 from page 15)

Batch	N	Presenting only five Interviews*	Cut score	R-square	Alpha	Exam Var
1	481	962	41.63	0.798	0.884	20.7%
2	147	294	41.66	0.852	0.908	13.9%
3	37	74	39.26	0.849	0.922	24.0%
Total	658	1330			0.891	
With 2 Interviews	7			SD	17.55	
				SEM	5.78	

* There were two sets of scores for each interview.

The metrics outlined in Table 5, also determine the cut-off point,⁴ for each batch (Batch 1 – 3 respectively, 41.63, 41.66, 39.26).

Inter-rater reliability

SWM23: With SWM there is no inter-rater issue since there is only one ‘rater’. The software uses the same algorithms and equations to calculate a suitability score for each respondent.

PVI: Inter-rater reliability is important for the PVI since the assessors are nominated by their respective employers, and offer two different perspectives on suitability for SW. The difference in perspective has the potential for an applicant being considered suitable based on insight into the course requirements, or insight into the challenges in SW practice, or vice versa. This puts a premium on inter-rater agreement since what one seeks is suitability for both the SW degree and practice. By summing the five scores⁵ on PVI questions from assessors one and two, for each

⁴ Pass mark regulations “No applicant should be awarded a place who scores below the pass mark”.

⁵ The Overall Score used to rank order applicants is the sum of the 10 scores provided (5 x 2) by A and B modified post hoc by the statistician for ‘technical reasons’ before forwarding to Admissions to allocate offers.

applicant and investigating using a Pearson Correlation we found that there was modest agreement on the suitability of applicants. More specifically *inter-rater correlation* showed a statistically significant linear relationship that was moderately positive ($r = 0.541$, $p = 0.000$).

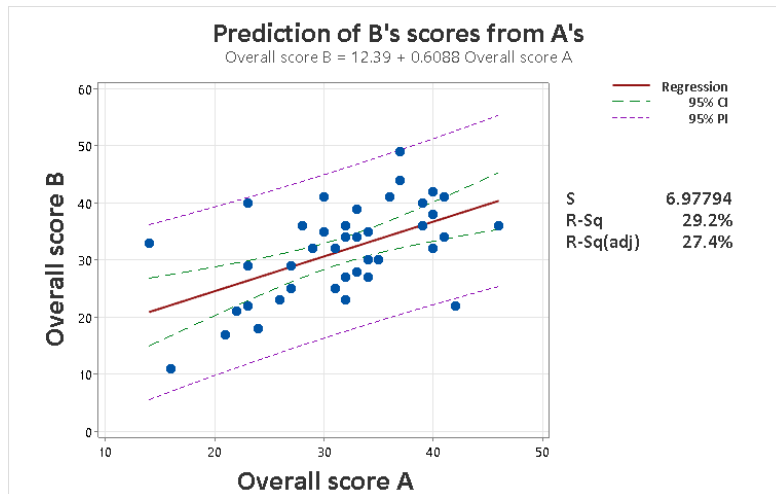


Figure 9: A positive linear relationship between A's and B's scoring of applicants

Using regression analysis we established a regression line as a visual indication of this linear relationship and the amount of variance in assessor two's scores that is explained by assessor one's scores ($S = 6.98$, $R\text{-square} = 29.2\%$) – a measure of the goodness of fit of the regression model and somewhat similar to the statistician's finding of variance for the three batches (exam var: 20.7%, 13.9%, 24.0%).

Spearman's rank correlation was computed to assess the relationship between the influential *Global Scores* (categories 1 - 5) provided by assessor one and those provided by assessor two. There was a positive correlation between the two variables, $r_s(39) = 0.508$, $p = 0.001$. There exists a moderate positive monotonic relationship between the global categories provided independently by the two assessors, that is as assessor one's ratings increased, assessor two's ratings tended to increase but not necessarily in a linear fashion. This was like the moderate positive relationship (except for the linearity) found to exist with the scoring of the five questions by assessors one and two.

In conclusion, since there is no subjective judgement involved in scoring, which is done automatically, and there is evidence of research underpinning the creation of

the SWM23 test, which was designed to measure and score the construct 'suitability for SW' the SWM23 test provides results with excellent test-retest, internal and inter-rater reliability and consistently reflects the reality of SW values held by applicants.

The PVI was not intended to be reused on the same individual (have test-retest reliability) but was designed to be a reliable assessment in the sense of being a dependable measure, and, while we have no direct evidence of test-retest reliability the quality of the assessment data does suggest dependability. We do have evidence of differences in inter-rater reliability and internal consistency, but these are differences in levels of high reliability and consistency. For both the SWM23 test and the PVI, reliability is high enough to ensure that the outcomes are dependable and can be used to make valid conclusions or decisions.

1c. Fairness

SWM23: We found SWM23 to meet APA standards for fairness (APA, 2014; 2020). In accordance with international standards, all participants receive comparable and equitable treatment during each phase of the testing or assessment process. Each participant receives standard instructions for use of the same user interface. Scoring is automated according to the applicant's response and not subjective.

PVI: We found the PVI is a fair method but with some conditional concerns. Although participants are given the same instructions, are asked similar questions with the same time to read and answer them, there was evidence of subjectivity in how assessor one and assessor two scored the responses to questions and in their Global Scores. It is anticipated that this variation was due to a balanced appraisal with differences of opinion and perspective. Given that there is a bank of trained assessors across academic institutions and SW agencies, who are each allocated 15-45 PVIs to mark, and they can independently assess their batch within a three-week window, an assessor may change their mode of working, possible unfair bias due to personal preferences must be mitigated using structural evaluation and multiple assessors. As we have seen this is done with the PVI and seems to be effective.

In addition to the definitions of each Global Score, the admissions committee may wish to provide more detailed guidance in the training to assist assessors with how the Global Score is used in decision-making. Furthermore, all applicants are not asked the same questions, as they are changed between batches to avoid the risk of them being shared⁶. Assessors, who are allocated a unique ID number, normally mark across several batches of interviews. This enables the statistician to look for trends by marker (i.e. hawks or doves).

The concern about fairness being mitigated arises from the number of ways personal preferences can emerge and their effect on fairness. For example, some applicants may be judged more on the assessor's subjective preferences or expectations, rather than the requirements of the degree or perceived competence required for SW practice. In addition, there may be unconscious bias towards certain groups in society, a drive to recruit more males, or people with more relevant experience, or favour shown to applicants they perceive as being like themselves. Such unfairness can only be mitigated if detected. Whereas using PVIs in tandem with a SWM23 test that is not affected by these preferences would help to counteract any concerns about subjective bias and unfairness in the selection process.

In conclusion, the empirical evidence supports the assertion that these two different measures of suitability for the SW degree and practice are valid, reliable, fair, and standardised. The data produced by the PVIs and SWM23 meet psychometric standards such as normal distributions means and standard deviations that help interpret an applicant's scores against the cohort of applicants. The distribution of score data may be normal, but there are distinct differences between the means and standard deviations indicating that they are not measuring the same variable. Despite this, analysis was able to establish and graph a moderate positive linear correlation between them. Factor analysis established that both the SWM23 test and the PVI were assessing a single factor. Since the test is designed to measure competence in one domain, private personal values and take on self and SW, and the interview to have experienced SWs assess an applicant's response to questions about SW, seeking outward and visible signs of potential competence, this is not a

⁶ In 2022 there was a choice of three questions addressing 'self-care' and four addressing 'diversity'.

surprise. We can with some confidence label the single factor as 'personal value based' in the SWM23 test and 'social work competency based' in the case of the PVI.

Both the SWM23 test and PVI measures are good, but differ in scoring and rating, yet they correlate moderately well and in a linear manner, leaving room for new knowledge about each applicant, if used in tandem. Combining the two perspectives offers a more rounded view, which is beneficial to universities and SW agencies when assessing suitability. Using both will promote fairness and equity of assessment of all applicants and facilitate effective evaluation.

Findings from Qualtrics survey complete by applicants.

As outlined earlier, the original sample consisted of 174 (26.44%) consenting participants from 658 applicants for the SW degree courses. Although 49 applicants completed the SWM test, only 46 of these could be analysed due to gross errors in how three participants completed the test, which suggested a lack of engagement.

In terms of the demographic characteristics of the 46 participants who completed both the test and the survey 39 (84.8%) were female, 6 (13%) were male and 1 (2.2%) preferred not to say. The age of participants ranged from 18-64 years, with 32.6% aged 18-24 years, 15% aged 25-34 years, 32.6% aged 35-44 years, 19.6% aged 45-64, and one respondent aged 55-64 years. In relation to the course applied for, 35 participants (76%) applied for the 3-year undergraduate route (UGR), two (4.3%) applied for the 2-year relevant graduate route (RGR) and the remaining 9 (19.6%) applied for both courses. In terms of which university assessed the PVI, over half of respondents (n=26; 56.5%) were assessed by UU, with the remainder assessed by QUB (n=14; 30.4%) or OU (n=6; 13%).

When asked *Was it easy to complete?* Participants responded as follows:

'Yes' (n=11; 24%), 'Mostly' (n=29; 63%) or 'No' (n=6; 13%).

When asked *Was it interesting to complete?* Participants responded as follows:

'Yes' (n=22; 48%); 'Mostly' (n=19; 41%) or 'No' (n=5; 11%). Based on Fisher's exact test ($p = .042$, two-tailed), there was a significant association between gender and whether respondents found the test interesting to complete, suggesting that a higher number of female respondents found SWM more interesting to complete than males.

When asked *Was it easy to understand the questions?* Participants responded as follows: 'Yes' (n=12; 26%); 'Mostly' (n=30; 65%) or 'No' (n=4; 9%). Based on Fisher's exact test ($p = .003$, two-tailed), there was a significant association between the course respondents applied for (UGR, RGR, both) and whether or not they found the questions easy to understand (Table 20).

Table 21. Was it easy to understand the questions on SWM23?

Course applied for	Yes	No	Mostly
Undergraduate route	7 (20.0%)	1 (2.9%)	27 (77.1%)
Relevant graduate route	2 (100%)	0 (0%)	0 (0%)
Both routes	3 (33.3%)	3 (33.3%)	3 (33.3%)

Note. Presented are within-course applied for percentages

When asked *Were you able to make quick intuitive response?* Results were as follows: 'Yes' (n=16; 35%); 'Usually' (n=21; 46%); 'Sometimes' (n=7; 15%) or 'No' (n=2; 4.3%). There was no significant association with any of the other variables.

When asked *Did you find it hard work to know how to respond?* Results were as follows: 'Yes' (n= 5; 11%); 'No' (n= 12; 26%); 'Usually' (n=2; 4.3%) or 'Sometimes' (n=27; 59%).

When asked *Did you think any of the values were not important?* All participants responded 'No'. when asked *Do you think any 'questions' made no sense?* The majority said 'No' (80.4%), with the remainder stating 'Yes' (19.6%).

Four respondents who answered 'Yes' elaborated as follows: One noted difficulty remembering what their favourite teacher thought as they attended school a long time ago. One could not understand how social workers could promote social justice, given the injustice in the real world. Another said they could have scored themselves at either end of the same scale for some questions, and the fourth did not know how a person s/he disliked would think.

When asked *Did you think any key social work values were missing?* The majority (n=40; 87%) said 'No' and the remainder stated 'Yes' (n=6; 13%). When prompted to state which values were missing, respondents indicated: confidentiality, autonomy, honesty, privacy, social justice, and safe/effective and high-quality care.

When asked *Did you feel you needed more time to complete it properly?* The majority stated 'No' (n=43; 93.5%) with the remaining 3 participants stating 'Yes'. When prompted, one indicated that they needed an additional 15 minutes, one completed it over two sittings (days) to enable them to read and fully understand the questions, and the other was unable to say, but stated it involved too much reading.

When asked *Did you find the test challenging to complete?* Responses were as follows: 'Yes' (n=6; 13%); 'No' (n=7; 15%) and 'A bit here and there' (n=33; 72%).

When asked *Overall, how did you find this experience compared to other selection processes you have experienced?* Responses are outlined in Table 21 as follows:

Table 21. Comparing this SWM23 experience to other selection processes.

Response	n (%)
Better	17 (37.0)
Worse	2 (4.3)
Easier	1 (2.2)
Harder	5 (10.9)
Different	15 (32.6)
Much the same	6 (13.0)
Total	46 (100)

When prompted for other comments, nine provided feedback as positive and negative experiences. Positive experiences were as follows:

“I enjoyed answering the questions mainly because it required some self-reflection on my part. It also may be highlighted some of my prejudice towards family and friends” ...

“I enjoyed participating in this kind of research” ...

“The test overall was very interesting in its use of different perspectives and getting into the minds of others” ...

“It would be very interested to receive the results of this study, including any identified limitations and a summary of the conclusions” ...

“I think this questionnaire is good to coincide with an applicant's interview score as I believe that someone's personal values, opinions and aspirations cannot always be gauged in an interview.”

In contrast, four participants were critical of the SWM test experience:

“I do prefer face to face interviews so I can really be myself. I am a people person, and I would like to show that side of me.”

“I think for some people it may be too repetitive and frustrating to complete”.

“It was very long, and I was struggling to understand some questions. I feel it would be good to add a ‘why’ onto them too and give the option to explain if you wanted to.”

“There are too many questions asking the same thing.”

In summary, most applicants found the test easy to use, easy to understand and interesting to complete. Over half stated that it was ‘sometimes’ hard work to know how to respond to questions, yet most were ‘usually’ or able to make quick intuitive responses.

For the minority of respondents who thought some questions made no sense, qualitative comments suggested this was linked to uncertainty about how to respond. In the instructions, applicants were advised to use the centre (zero) to deal with such uncertainty. Only six respondents thought a key social work value was missing from the SWM test, most of which were included. Most participants were able to complete the test within the recommended time of 45-60 minutes.

Interestingly, a large percentage of participants (37%) thought SWM23 was a better way for assessing the suitability of applicants to SW training, with one suggesting it would be good to use SWM alongside the PVI and another recognising the importance of self-reflection in SWM.

Conclusion and Recommendations

Fitness for purpose of SWM23 and PVI as selection tools

When two different measures of the same variable correlate to a moderate degree, as they do here, it suggests that they share some common information about the variable, but each measure also provides a significant amount of unique information. The fitness for purpose of both measures could be improved by using them in tandem, as it would give a better overall understanding of a person's suitability for SW than either test alone, because they capture different aspects or manifestations of the suitability construct (personal values and key competencies).

Ten years ago, Croisdale-Appleby (2014) recommended the introduction of values-based recruitment methods to improve the selection process. Furthermore, findings from the review of complaints to the NISSC (Hayes, 2018) highlighted evidence of SWs being dishonest, disrespectful and discriminatory towards service users. Research into reasons why SW students fail practice placements across Ireland (Roulston et al., 2021) indicated that several students failed due to values related issues such as being unable or unwilling to follow guidance, professional misconduct, poor time management, breaching data protection, or being oppressive towards service users. Using SWM23 to assess the personal and professional values of applicants would add another element to identifying those who may be unsuitable for professional training, prevent them entering the social work profession and minimise the risk of bringing the SW profession into disrepute.

Practical usefulness of SWM and PVI

Empirical results indicate that both SWM23 and PVI are valid, reliable, and fair measures of different aspects of suitability, which can effectively rank order and classify applicants. Any uncertainty about the fairness of how applicants are assessed at PVI can be reduced by using the measures in tandem. Feedback from participants indicated that the SWM23 test was easy to understand and interesting to complete, which mirrors previous feedback on SWM (Roulston et al., 2022).

Admissions staff have indicated difficulties in selecting the next suitable applicant, when several have been awarded the same score, and they do not have sufficient places to offer all applicants the same score a place. Having applicants rank ordered does assist with this, unless more than one student is awarded the same PVI score. Having rank ordered SWM23 scores to use for decision-making would enhance the

rigour of the selection process, and ensure places on the course are allocated to applicants with the strongest core values.

Strengths and limitations of SWM23 and the PVI

SWM23 is concerned with identifying the applicant's core values and the strength of feeling underpinning them, how they make sense of themselves, other people, and society in general in the context of social work. The main limitation is that it is self-report, self's perspective on, and appraisal of, self and the world, although it is important for a university or prospective employer to have a sense of what that is.

The PVI is concerned with understanding the applicant as a person and having a trained assessor balance what they hear and see with what they perceive is required of a social work student at university and in practice. Although the interview does touch on core values and beliefs, it does not focus on exploring them or trying to identify the private, often subconscious unspoken, values, attitudes and feelings that will constrain and shape the applicant's behaviour in the real world of social work practice. Furthermore, as the independent assessors come from different perspectives and have different levels of experience working with students at the outset of their professional training, they can award different scores for an applicant in relation to the same question or competency, as illustrated above. This difference of opinion regarding scores for individual questions is then disguised by the calculation of the overall (ranking) score. Therefore, the differences between the SWM23 and the PVI, the overlap and positive correlation in the measures suggests they are best used in tandem.

References

American Psychological Association. (2020). Publication manual of the American Psychological Association (7th ed.). <https://doi.org/10.1037/0000165-000>

APA Publication Manual 7th Edition (2020). <https://apastyle.apa.org/style-grammar-guidelines/references>

American Educational Research Association, [2014] Standards for educational and psychological testing / American Educational Research Association, American Psychological Association, National Council on Measurement in Education.

Baginsky, M. (2013). *Retaining experienced social workers in children's services: The challenge facing local authorities in England*. London: Department for Education.

Bhandari, P. (2022). *What Is Interval Data? Examples & Definition*. Scribbr. Retrieved 24 July 2023, from <https://www.scribbr.co.uk/stats/interval-data-meaning/>

Cleland, J. A., Dowell, J., McLachlan, J., Nicholson, S., & Patterson, F. (2012). Identifying best practice in the selection of medical students: Literature review and interview survey. General Medical Council. Available at: <https://www.gmc-uk.org/-/media/gmc-site/about/identifyingbestpracticeintheselectionofmedicalstudentspdf51119804.pdf>

Clinedinst, M. (2019). State of college admissions. The National Association for College Admission Counselling. https://www.nacacnet.org/globalassets/documents/publications/research/2018_soca/soca2019_all.pdf

Cree, V., Morrison, F., Clapton, G., Levy, S., & Ingram, R. (2018). Selecting social work students: Lessons from research in Scotland. *Social Work Education*, 37(4), 490–506. <https://doi.org/10.1080/02615479.2017.1423050>

Croisdale-Appleby, D. (2014). Re-visioning social work education: An independent review. https://www.basw.co.uk/system/files/resources/basw_22718-9_0.pdf

Cronbach, L.J. and Meehl, P.E. (1955). "Construct validity in psychological tests". *Psychological Bulletin*. 52 (4): 281–302. <https://doi.org/10.1037/h0040957>

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co.

Ferguson, E., Sanders, A., O'Hehir, F., & James, D. (2000). Predictive validity of personal statements and the role of the five-factor model of personality in relation to medical training. *Journal of Occupational and Organisational Psychology*, 73(3), 321–344. <https://doi.org/10.1348/096317900167056>

Hackett, S., Kuronen, M., Matthies, A.-L., & Kresal, B. (2003). The motivation, professional development and identity of social work students in four European countries. *European Journal of Social Work*, 6(2), 163–178. <https://doi.org/10.1080/1369145032000144421>

Hayes, D. (2018). Relationships matter: An analysis of complaints about social workers to the Northern Ireland social care council and the patient and client council. Northern Ireland Social Care Council. <https://www.basw.co.uk/resources/relationships-matter-analysis-complaints-about-social-workers-northern-ireland-social-care>

Klieger, D. M., Belur, V. and Kotloff, L. J. (2017). Perceptions and uses of GRE® scores after the launch of the GRE® revised General Test in August 2011. ETS Research Report Series (pp. 1–49). <https://doi.org/10.1002/ets2.12130>

Kuncel, N., Tran, K. and Zhang, S. (2020). Measuring Student Character: Modernizing Predictors of Academic Success. In M. Oliveri & C. Wendler (Eds.), *Higher Education Admissions Practices: An International Perspective* (Educational

- and Psychological Testing in a Global Context (pp. 276–302). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108559607.016>
- McCartan, C., Byrne, J., Campbell, J., Coogan, D., Davidson, G., Hayes, D., Kirwan, G., MacDonald, M., McCaughren, S., McFadden, P., McGregor, C., Montgomery, L., Pentaraki, M., Roddy, S., Roulston, A., Shore C. and Wilson, E. (2022) Social work students on the island of Ireland: a cross sectional survey, *Social Work Education*, 41:2, 228-247, <https://doi.org/10.1080/02615479.2020.1832060>
- McFadden, P. (2018). Two sides of one coin? Relationships Build resilience or contribute to burnout in child protection social work: Shared perspectives from leavers and stayers in Northern Ireland. *International Social Work*, 63(2), 164–176. <https://doi.org/10.1177/0020872818788393>
- McFadden, P., Mallet, J., Campbell, A., & Taylor, B. (2019). Explaining self-reported resilience in child protection social work: The role of organisational factors, demographic information and job characteristics. *British Journal of Social Work*, 49(1), 198–216. <https://doi.org/10.1093/bjsw/bcy015>
- McNeill, C., Traynor, M., Erskine, A., & Ellis, R. (2018). Developing nurse match: A selection tool for evoking and scoring an applicant's nursing values and attributes. *Nursing Open*, 6(1), 59–71. <https://doi.org/10.1002/nop2.183>
- Messick, S. (1995). The once and future issues of validity: Assessing the meaning and consequences of measurement. In *Test validity*. <https://doi.org/10.1002/j.2330-8516.1986.tb00185.x>
- Patterson, F., Prescott-Clements, L., Zibarras, L., Edwards, H., Kerrin, M., & Cousans, F. (2016). Recruiting for values in healthcare: A preliminary review of the evidence. *Advanced Health Science Education*, 21(4), 859–891. <https://doi.org/10.1007/s10459-014-9579-4>
- Patterson, F., Roberts, C., Hanson, M. D., Hampe, W. E., Ponnampereuma, K., Magzoub, G., Cleland, A., & Tekian, J. (2018). Ottawa consensus statement:

Selection and recruitment to the healthcare professions. *Medical Teacher*, 40(11), 1091–1101. 2018. <https://doi.org/10.1080/0142159X.2018.1498589>

Roulston, A., Cleak, H., Hayes, D., McFadden, P., O'Connor, E., & Shore, E. (2021). To fail or not to fail: Enhancing our understanding of why social work students failed practice placements (2015-2019). *Social Work Education: The International Journal* <https://doi.org/10.1080/025615479.2021.1973991>

Roulston, A., McNeill, C., Hayes, D., & Traynor, M. (2022). Evaluating shortlisting methods for admission into the social work degree: personal statement versus psychological test. *Social Work Education*, 1–21. <https://doi.org/10.1080/02615479.2022.2146668>

Ryan, M., Cleak, H., & McCormack, J. (2006). Student performance in field education placements: The findings of a 6-year Australian study of admissions data. *Journal of Social Work Education*, 42(1), 67–84. <https://doi.org/10.5175/JSWE.2006.200303106>

Stevens, M., Moriarty, J., Manthorpe, J., Hussein, S., Sharpe, E., Orme, J., Mcyntyre, G., Cavanagh, K., Green-Lister, P. and Crisp, B. R. (2012). Helping others or a rewarding career? Investigating student motivations to train as social workers in England. *Journal of Social Work*, 12(1), 16–36. <https://doi.org/10.1177/1468017310380085>

Stratton, K. (2000). Before and after science: Admissions process to Australian university social work courses. *Australian Social Work*, 53(3), 29–35. <https://doi.org/10.1080/03124070008414313>

Traynor, M., Barr, O., McNeill, C., Erskine, A., & Ellis, R. (2019). Burdett trust for nursing final report 'what matters to patients': Identifying applicants to nursing who have the personal values required to build a skilled and competent workforce and lead the nursing profession in Northern Ireland. Executive Summary A. Belfast. https://www.qub.ac.uk/schools/SchoolofNursingandMidwifery/FileStore/Filetoupload_945100.en.pdf

Turney, S. (2023). Skewness | Definition, Examples & Formula. Scribbr, www.scribbr.co.uk

Wilson, G. and McCrystal, P. (2007) Motivations and Career Aspirations of MSW Students in Northern Ireland, *Social Work Education*, 26:1, 35-52. <https://doi.org/10.1080/02615470601036534>

Woo, T., Ho, R., Tang, A., & Tam, W. (2020). Global prevalence of burnout symptoms among nurses: A systematic review and meta-analysis. *Journal of Psychiatric Research*, 123, pp.9-20. <https://doi.org/10.1016/j.jpsychires.2019.12.015>