

DEVELOPMENT AND VALIDATION OF AN APPROACH FOR SELF-ASSESSMENT OF COMMUNICATION SKILLS IN CLINICAL DENTISTRY:

A mixed methods evaluation in Syria

Ghaith Alfakhry

Department of Education

University of Oxford

St. Antony's College

Supervisors:
Associate Professor Ariel Lindorff
& Dr. Abdul Karim Ismail

Thesis submitted to the University of Oxford in partial fulfilment of the requirements for the Degree of MSc in Education (Research Design and Methodology).

August, (2024)

ABSTRACT

Communication skills are universally recognised as a core competency in clinical education, yet in many Arab countries, including Syria, these skills are not formally integrated into the curriculum. This study aimed to develop and validate a self-assessment approach for communication skills, incorporating patient assessments for corrective feedback, at Damascus University Faculty of Dental Medicine, Syria. The study investigated this assessment approach's reliability, validity, and educational impact.

Using the well-established Calgary Cambridge Guide, an assessment instrument was developed, translated and cross-culturally validated for the Syrian Arabic context. The assessment approach comprised six consultation sessions with six patients in clinical settings. After each consultation, students and patients completed an assessment form. Fifty-four final-year dental students completed all six sessions. A sequential mixed methods design was employed, beginning with a quantitative phase that included generalizability (G) and decision studies to assess reliability and exploratory factor analysis (EFA) to evaluate structural validity. In the qualitative phase, semi-structured interviews were conducted with a purposive subsample (n=12) from the quantitative strand. The thematic analysis explored the educational impact and factors influencing self-assessment.

Findings showed a G-coefficient of 0.93 for self-assessment and 0.66 for patient assessment. Student and rater effects are conflated in self-assessment G-study, so the coefficient needs to be interpreted cautiously. EFA indicated a one-factor solution explaining 52.1% of the variance. There was a significant improvement in patient assessments over time, with a large effect size ($P<0.001$, Partial $\eta^2=0.184$). Qualitative results supported these findings, revealing increased self-awareness and self-control due to self-assessment. Subjective self-assessment practices, patient-dentist dynamics, and self-reflection all influenced self-assessment scoring. Focusing on a context that lacks a formal communication skills curriculum, this rigorous mixed methods research makes an original and substantive contribution to the field. One original contribution of this study is the in-depth exploration of meanings and intentions underlying self-assessment results that most studies in the field overlook. Although the findings should be generalised cautiously, this research provides valuable recommendations for students, clinicians, and clinical educators on optimising self-assessment practice and using its data to facilitate communication performance improvement.

Acknowledgement

I would like to express my deepest gratitude to my supervisor, Professor Ariel Lindorff, who has been nothing but supportive and fair from my MSc application process to the moment I submitted my dissertation. Her expertise, critical insights and feedback were a cornerstone in the success of this work. Working under her supervision has been a privilege, and I am immensely grateful for everything she taught me throughout my MSc journey.

I would also like to thank Dr. Abd Alkarim Ismail, my co-supervisor, for his invaluable contribution and feedback on various stages of the research project, including design, analysis, and presentation. His creative ideas and insightful, detailed guidance have left a significant mark on this project.

My sincere thanks go to Dr. Osama Alrikabi, who provided essential support in data collection. I also wish to acknowledge the patient participants for their commitment and the student dentists for their participation in this study. Without their involvement, this research would not have been possible.

I would like to acknowledge the generous support of the Chevening Scholarship, which allowed me to pursue a master's degree in the UK and conduct this research.

To my partner, Rama Kodmani, for her patience and support. I am truly fortunate to have such an amazing, caring and loving partner by my side to keep me going throughout the lows of this journey.

TABLE OF CONTENTS

ABSTRACT ..... **ii**

TABLE OF CONTENTS ..... **iv**

LIST OF TABLES ..... **viii**

LIST OF FIGURES ..... **x**

LIST OF ABBREVIATIONS ..... **xi**

1. INTRODUCTION **1**

1.1 CHAPTER GUIDE 1

1.2 COMMUNICATION SKILLS IN MEDICAL EDUCATION: AN INTRODUCTION 1

1.3 CONTEXT FOR THE CURRENT STUDY: DENTAL EDUCATION IN WAR-TORN SYRIA 3

1.4 STUDY RATIONALE AND AIMS 4

1.5 STRUCTURE OF THE THESIS 4

2. LITERATURE REVIEW **5**

2.1 CHAPTER GUIDE 5

2.2 CONCEPTUALIZATION OF COMMUNICATION SKILLS 5

2.3 ASSESSMENT OF COMMUNICATION SKILLS 6

2.4 SELF-ASSESSMENT 8

2.4.1. Conceptualization of self-assessment 8

2.4.2 Perspectives and debates in self-assessment 10

2.4.3 Self-assessment of communication skills 12

2.5 WHAT DOES VALIDITY MEAN IN THE CONTEXT OF SELF-ASSESSMENT OF COMMUNICATION SKILLS? 13

3. MATERIALS & METHODS **15**

3.1 CHAPTER GUIDE 15

3.2 RESEARCH AIMS AND QUESTIONS 15

3.3 MIXED METHODS RESEARCH DESIGN....16
3.3.1 Remarks on methods integration....19
3.4 ASSESSMENT APPROACH DEVELOPMENT....19
3.4.1 Assessment aim....19
3.4.2 Assessment instrument development....19
3.4.3 Translation and cross-cultural validation into Arabic....21
3.4.4 Pilot testing of the translated instrument: Clarity and relevance....22
3.4.5 Final assessment instrument....22
3.4.6 Assessment approach: procedure and patient assessors....23
3.5 QUANTITATIVE STRAND: SETTINGS, SAMPLE, AND METHODS....24
3.5.1 Setting and student participants....24
3.5.2 Data collection....25
3.5.3 Missing data and outliers....26
3.5.4 Data Analysis....26
3.6 QUALITATIVE STRAND: SETTINGS, SAMPLE, AND METHODS....30
3.6.1 Setting and participants....30
3.6.2 Data generation and analysis....31
3.7 REFLEXIVITY AND POSITIONALITY....33
3.7.1 Theoretical stance....33
3.7.2 Personal and professional background....33
3.8 COMMENTS ON RIGOUR....34
3.9 ETHICAL CONSIDERATIONS....35
3.10 ACCESS TO RESEARCH DATA AND MATERIALS....37
4. RESULTS....38
4.1 CHAPTER GUIDE....38
4.2 RELIABILITY....38
4.2.1 Reliability of self-assessment....38

4.2.2 Reliability of patients' assessment 39
4.3 STRUCTURAL VALIDITY 42
4.4 FURTHER QUANTITATIVE ANALYSIS 45
4.5 IMPACT OF THE ASSESSMENT APPROACH 49
4.5.1 Self-control 49
4.5.2 Now I see it! 51
4.5.3 Self-assessment versus patient assessment 52
4.6 SELF-ASSESSMENT UNDER THE MICROSCOPE: EXPLORING JUSTIFICATIONS 53
4.6.1 Subjective use of the assessment form 53
4.6.2 Patient-dentist dynamics 55
4.6.3 Self-reflection 56
5. DISCUSSION AND CONCLUSIONS 58
5.1 CHAPTER GUIDE 58
5.2 VALIDITY OF THE ASSESSMENT APPROACH 58
5.2.1 Content and ecological validity 58
5.2.2 Reliability of the assessment approach 58
5.2.3 Structural validity of the assessment tool 61
5.2.4 Validity as impact: Integrated conclusions from quantitative and qualitative strands 62
5.3 WHAT DO THESE SELF-GIVEN SCORES MEAN? 64
5.4 CONTRIBUTION TO THE RELEVANT ACADEMIC LITERATURE 66
5.5 STRENGTHS AND LIMITATIONS 66
5.6 DIRECTIONS FOR FUTURE RESEARCH 67
5.7 IMPLICATIONS FOR EDUCATIONAL PRACTICE 68
REFERENCES 69
APPENDICES 80
APPENDIX I: First version of the developed assessment instrument 80

APPENDIX II: Pilot testing participants' characteristics and findings.....81
APPENDIX III: Assessment instrument (English and Arabic versions) 83
Appendix IV: G*Power configuration and output 91
APPENDIX V: Missing data and outliers.....92
APPENDIX VI: Example R software code for performing Generalisability and Decision studies on self-assessment data.....94
APPENDIX VII: Interview protocol (English version) and modifications. 95
APPENDIX VIII: Ethical approval, amendments, participants facing materials, and consent forms 98
APPENDIX IX: Rotated factor loadings for the first two factors with direct oblimin rotation 114
APPENDIX X: Descriptive statistics, skewness and kurtosis, and tests of normality for aggregated 30 items of self-assessment and patient assessment (ssum30=aggregated SA, psum30=aggregated PA). 115
Appendix XI: Mixed repeated measures ANOVA for SA and PA for the defined three groups based on patient assessment tiers. 116

LIST OF TABLES

Table 1. Student participants' characteristics (n=54) 25
Table 2. Score ranges and number of participants based on their tier on self-assessment and patient assessment data. 30
Table 3. Distribution of self-assessment high, mid, and low groups across the three counterpart patient tiers.....31
Table 4. G-study results for the p (student) × i (item) × c (case) design.....38
Table 5. D-study results for the p (student) × i (item) × c (case) design, showing the variance for each component and interaction. 39
Table 6. G-study results for the p (student) × i (item) × r (rater) design. 40
Table 7. D-study results for the p (student) × i (item) × r (rater) design, showing the variance for each component and interaction. 40
Table 8. Comparison of G-studies between standardized patients and real patients (Numbers: person=1, items=30, rater=3,). 41
Table 9. Initial Eigenvalues and total variance explained by extracted factors. 43
Table 10. Descriptive statistics and factor matrix for 324 filled forms (54 participants x 6 examinations). 44
Table 11. Descriptive statistics for the aggregated self-assessment and patient assessment as per patient ID with One-way ANOVA test. 46
Table 12. Mixed repeated measures ANOVA statistics for aggregated self-assessment and patient assessment as per patient order. 46
Table 13. Pearson correlation and confidence intervals between Self-assessment, patient assessment, university grade, initial global SA (assessed before approaching patients) and the global SA item (assessed during sessions for each patient) 48
Table 14. Interview student participants' characteristics and mean self- and patient assessment for the six encounters. 49
Table 15. Summary of previous G-studies in the assessment of communication skills....61
Table 16. The first version of the developed assessment instrument. 80
Table 17. Participants' characteristics for the pilot test based on self-report. 81
Table 18. Pilot testing results of the translated instrument by patients, clinical dental students and experts. 82

Table 19. Item-non-response analysis for both self-assessment and patient assessment data. 92
Table 20. Mean and standard deviation before and after deleting outliers for the self-assessment and patient assessment dataset. 93
Table 21. A summary of the amendments made to the CUREC 1A form 99
Table 22. Mixed repeated measures ANOVA for self-assessment and patient assessment for the defined three groups based on Patient assessment tiers. 116

LIST OF FIGURES

Figure 1. Conceptual framework for the clinician- and patient-related factors that affect clinician-patient interaction in the medical consultation. 2
Figure 2. The Calgary-Cambridge guide framework. 6
Figure 3. Self-regulated learning cyclical model adapted from Zimmerman (2000) and Artino (2012). 11
Figure 4. Diagram showing the specific research process and interactivity between QUAN and qual strands. A slight variation of Morse conventions was followed in the presentation (Morse, 1991). QUAN=quantitative; capital letters are used to show priority. qual=qualitative. 18
Figure 5.Diagram showing the adapted framework from Calgary-Cambridge Guide used in this study. 21
Figure 6 Two Venn diagrams for self-assessment (right) and patient assessment (left) showing a fully crossed two-facet design: (p × i × c) and (p × i × r). 27
Figure 7. Generalisability coefficients for different numbers of patient assessors in the patient assessment data. 41
Figure 8. Generalisability coefficients for different numbers of patients (cases for self-assessment and raters for patient assessment) across the three raters (self-assessment [SA], standardized [SP] and real patient [RP] assessments)..... 42
Figure 9. Scree plot for the extracted factors. 43
Figure 10. Aggregated patient assessment and self-assessment as per the chronological order of patient consultation (Time) 47
Figure 11. Patient assessment is represented as a line that is intersected with a self-assessment line. The blue line is when self-assessment is equal to patient assessment. The areas above the blue line (A and D)represent those with higher SA and lower PA. Areas below the blue line (B and C) represent students with lower SA and higher PA..... 64
Figure 12 G*Power configuration to calculate the sample size 91

LIST OF ABBREVIATIONS

CCG: Calgary Cambridge Guide

CFA: Confirmatory factor analysis

CUREC: Central University Research Ethics Committee

EFA: Exploratory factor analysis

G-coefficient: Generalizability coefficient

MM: Mixed methods

PA: Patient assessment

RP: Real patients

SA: Self-assessment

SEM: Standard error of measurement

SL: Source language

SP: Standardized patients

SRL: Self-regulated learning

TL: Target language

η^2: Eta square

1. INTRODUCTION

1.1 CHAPTER GUIDE

This chapter is intended to:

- Introduce the field and background for the study
- Describe the definitions and significance of communication skills
- Describe the context of the study
- Present the rationale and purpose of the study
- Outline the structure of the dissertation

1.2 COMMUNICATION SKILLS IN MEDICAL EDUCATION: AN INTRODUCTION

Communication skills are universally recognised as a fundamental competency by prominent regulatory and medical educational bodies, including the General Dental Council (GDC) and the General Medical Council (GMC) in the UK, the Accreditation Council for Graduate Medical Education (ACGME) in the US, and the World Federation for Medical Education (WFME) (ACGME, 2023; GDC, 2015; GMC, 2023; WFME, 2015). Clinicians engage with hundreds of thousands of patients over a lifetime; thus, communication and interpersonal skills make up the very fabric of clinical practice (Cushing, 1996). Good clinician communication positively affects patient satisfaction, and research shows that it is associated with improved healthcare outcomes (Fox, 2010), better adherence (Shigli & Awinashe, 2010), and a lower chance of being complained about or, more seriously, litigated against by patients (O'Hara et al., 2018). A 2022 report by the National Health Service in England showed that most written complaints from patients were related to clinician-patient communication. (NHS, 2022).

The concept of good communication for clinicians is difficult to unravel and is very context-, case- and patient-bound. Skilled communication transcends the verbal aspect to include implicit attributes such as empathy, sensitivity, engagement, attitude, respect, and authenticity, which can be communicated through non-verbal means (Verheijden et al., 2023). Clinicians need to communicate appropriately with patients and their families, friends, and other concerned parties. Many factors play a role in clinician-patient interaction, including patient- and clinician-related factors and their interaction, illustrated

in **Figure 1** (Ashton et al., 2003; Bertakis, 2009; Paasche-Orlow et al., 2005). Age, gender, race, socioeconomic status, attitudes, beliefs and communication skills are factors that exist at the patient and the clinician level that can affect the quality of interaction (Ashton et al., 2003; Bertakis, 2009). As for the interaction between patient- and clinician-related factors, studies have shown that gender-concordant visits between clinician-patient dyads could play a positive role in the interaction happening at the medical visit (Bertakis, 2009). Another study showed that racial concordance between the patient and the clinician is associated with better healthcare use in minority populations (Jetty et al., 2022).

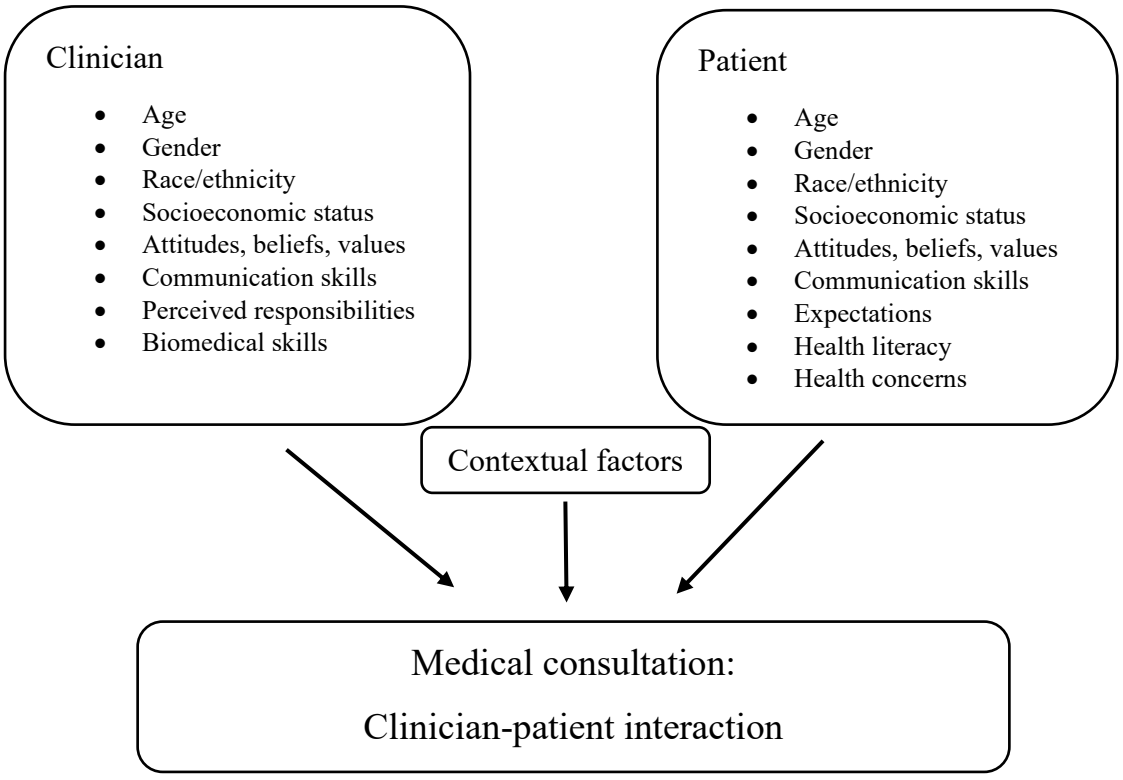


Figure 1. Conceptual framework for the clinician- and patient-related factors that affect clinician-patient interaction in the medical consultation.

To add to the complexity, patients' cultural and societal backgrounds might differ from the clinician's, necessitating a nuanced consideration of these aspects during interaction (Claramita et al., 2013). What is more, is that the biomedical and communication tasks can be interdependent (Keller & Carroll, 1994); for instance, diagnosing a patient relies heavily on the clinician's ability to gather detailed patient history; alternatively, a patient

with severe illness and health concerns could demand higher sensitivity and empathy compared to a patient with less severe needs (Bensing et al., 2003). This makes teaching and assessing communication skills for healthcare professions difficult due to the conceptual, operational and contextual complexity of the skills and behaviours (Cushing, 1996; Norman, 2005).

1.3 CONTEXT FOR THE CURRENT STUDY: DENTAL EDUCATION IN WAR-TORN SYRIA

In Syria, public universities such as Damascus University offer free education for Syrian nationals. Students enrol in dental school usually at the age of 18 after taking the standardised 12th-grade exam (Baccalaureate examination), which, according to its score, students can apply to study at university. Medical, dental and pharmacy schools are consistently the most competitive, usually requiring a minimum cumulative score of approximately >95% on the exam. There are no entry examinations to dental schools. Even though Damascus University dental school is considered the best in the country, it was significantly affected by the ongoing conflict in the country, which started in 2011(Shehada et al., 2023). Some well-documented issues are a lack of resources, a high student/faculty ratio, and outdated lecture-based and teacher-centred approaches. (Alfakhry, Mustafa, AlMukhallalati, et al., 2023; Dashash, 2019; Shehada et al., 2023). These issues have worsened over the years of conflict and the economic collapse of the national currency.

The dental undergraduate curriculum at Damascus University is divided into three pre-clinical years and two clinical years. To graduate, students must successfully complete all their modules and take the Standardized National Dental Examination as a final pre-requisite for graduation. In the last two clinical years, dental students have been required to treat a wide variety of complex clinical cases, including extraction, restorative, removable, and fixed prosthodontics, pedodontics, and endodontics. The number of cases per module varies but nonetheless is considered demanding, reaching up to 12-15 treated cases in some modules. Communication skills are not taught at the dental school; some informal notes are made on how to treat patients in the clinics by the supervisors. However, these are generally broad notes to respect and be kind to the patients. Generally speaking, students choose their way of communicating with the patients, some of which can be respectful while others can be outright inconsiderate (Alfakhry, Mustafa, Jazayerli,

et al., 2023). Thus, a specific focus on communication skills development could be very much needed in this setting.

1.4 STUDY RATIONALE AND AIMS

In an effort to address this critical gap, this study aims to develop and validate an instrument and an approach for self-assessment of communication skills in patient consultations, targeting final-year dental students at Damascus University. The approach integrates patient assessments to provide feedback and complement students' self-assessments.

1.5 STRUCTURE OF THE THESIS

This chapter overviews the study's background, context, rationale, and aims. Chapter two reviews the literature on communication skills and self-assessment in clinical education, including key debates. Chapter three outlines the research design and methodology, starting with the study's aims and research questions, followed by an explanation of the mixed methods design and the development of the assessment method. It details the quantitative and qualitative approaches and concludes with a discussion of rigour and ethical considerations. Chapter four presents the findings, while chapter five discusses these findings within the context of relevant literature.

2. LITERATURE REVIEW

2.1 CHAPTER GUIDE

This chapter is intended to provide the reader with an overview of theory and research in two areas of medical education: communication skills and self-assessment. Validity literature in educational assessment will briefly be discussed in the context of this study.

2.2 CONCEPTUALIZATION OF COMMUNICATION SKILLS

Kurtz et al. (2017) conceptualisation of communication skills is one of the highly cited, theoretically supported and popularly used in communication teaching and assessment programs for health professionals (Kurtz et al., 2017; Kurtz et al., 2003; Kurtz & Silverman, 1996; Silverman et al., 2016). In this conceptualization, communication skills are defined by three broad skills: content skills (what is communicated), process skills (how it is communicated) and perceptual skills (self-awareness about feelings and thoughts). Content skills pertain to what clinicians communicate, encompassing the materials of their interactions, including questions and answers, the data they collect, analyse, and the conclusions they produce and the treatments they discuss. Process skills are concerned with *how* clinicians communicate, i.e. their method/approach/style of communication, which includes how they conduct history taking, examine the patient, provide information, structure their communication, develop rapport/relationship with the patient; the clinician's questioning skills, language skills, and non-verbal skills. Lastly, perceptual skills are about internal decision-making, clinical reasoning, and problem-solving skills; the attitudes, beliefs, intentions, and values; their self-awareness of feelings and thinking towards the patient; their confidence, self-identification and reflections about their own biases and positionality (Kurtz et al., 2017, p. 32). Content, process and perceptual skills are interlinked and cannot be considered in isolation when discussing communication skills. These skills provide a broad framework; however, more detailed skills that ground clinician-patient communication skills are provided in the form of the Calgary-Cambridge Guide (CCG). This guide was first introduced in 1998 and then been updated in 2003 (Kurtz et al., 1998; Kurtz et al., 2003). The CCG has 71 items that fall under six broad communication tasks that are chronologically ordered, with two additional tasks that occur continuously throughout the consultation (**Figure 2**) (Kurtz et

al., 2017, p. 41). This detailed framework underscores the multifaceted nature of communication skills in healthcare, setting the stage for a nuanced approach to assessing these competencies with clear, actionable criteria. By grounding the study in this well-established framework, the research ensures that the assessment instrument used is aligned with widely accepted standards, thereby enhancing the validity and generalizability of the findings.

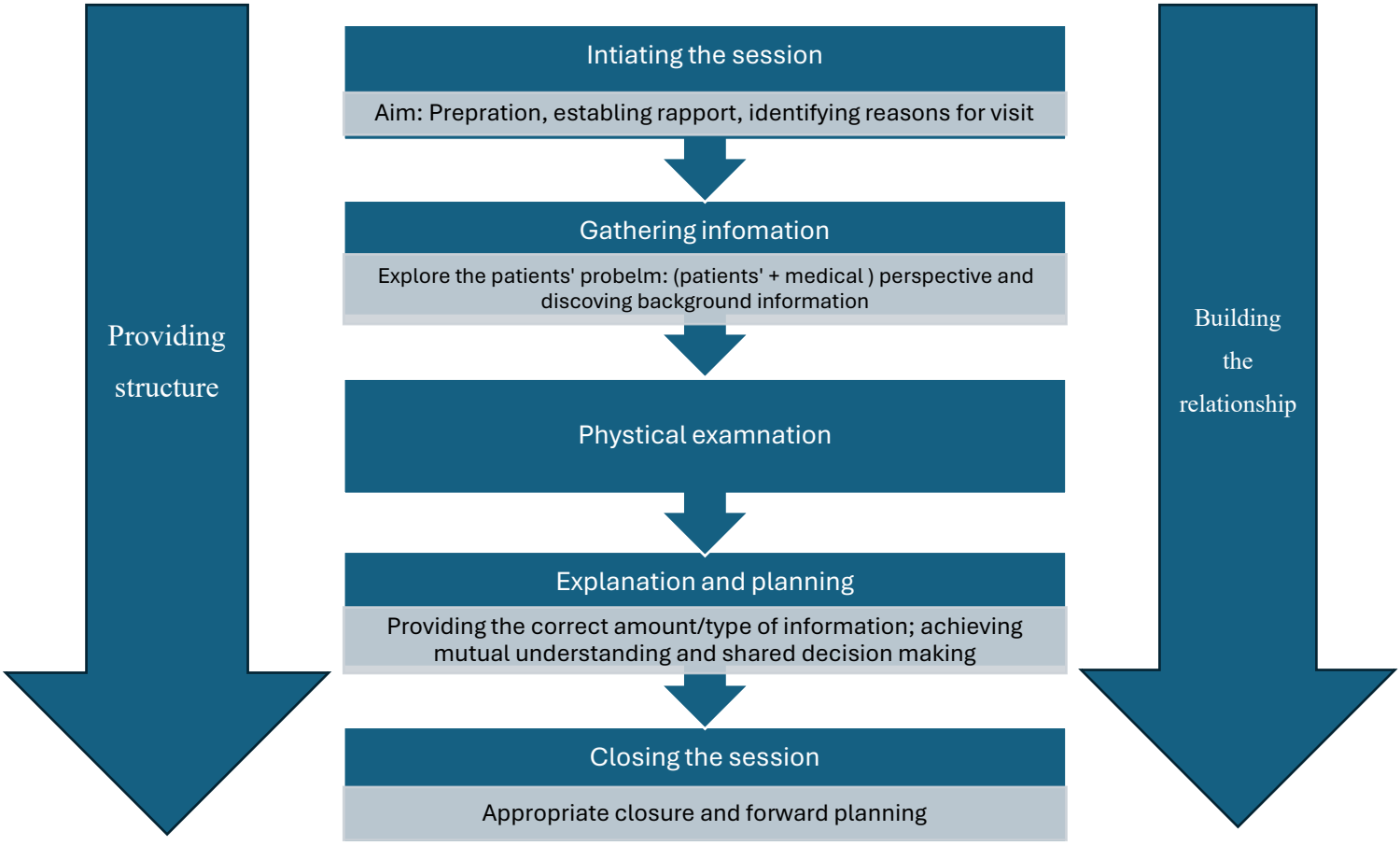


Figure 2. The Calgary-Cambridge guide framework.

2.3 ASSESSMENT OF COMMUNICATION SKILLS

The importance of assessment in an education system transcends stages and disciplines. While there has been a push towards the "*assessment for learning*" movement, the dominant paradigm remains "*assessment drives learning*"; that is, examinations and assessments dictate what should be taught and learnt by students (Patel et al., 2018).

Therefore, ensuring that the assessment goals are aligned with curriculum goals is of utmost importance. If students can pass an exam without achieving the desired skills and level of performance, the assessment fails to serve its intended purpose (Wass et al., 2001).

Assessment in medical education encompasses various constructs, with some assessments serving certification and licensure purposes, known as summative assessments, while others are designed to provide students with feedback for their ongoing development, referred to as formative assessments (Schuwirth & Van der Vleuten, 2018). One definition that I will adopt in this dissertation is given by Schuwirth and Van der Vleuten (2018, p. 277) defining assessment as "*any purported and formal action to obtain information about the competence and performance of a candidate.*."

Any assessment should start with a clear blueprint of the object of assessment i.e. planning the test against the competencies essential for successful practice (Wass et al., 2001). An example blueprint for communication skills assessment can be the conceptual framework illustrated in the previous section. In communication skills assessment, van der Vleuten et al. (2019) outlined a number of principles that need to be considered; the assessment needs to:

1. Longitudinal
2. Authentic as possible
3. Provide meaningful feedback
4. Enable valid fail/pass decision making

Various methods can be used to address these purposes. However, due to the behavioural nature of communication skills, direct observation is the preferred method (van der Vleuten et al., 2019). Feedback generated from direct observation also seems to be more credible and meaningful to learners (Sargeant et al., 2010). According to the second principle, direct observation is most authentic when conducted in real clinical settings. Govaerts et al. (2007) highlighted the benefits of in-vivo performance assessment, which combines knowledge, technical and cognitive skills, and professional behaviour, thus breaking the discordance between assessment and experienced clinical practice, ultimately making the assessment more meaningful to learners (Sargeant et al., 2010).

The assessors can be external examiners, clinicians, other health professionals, real patients or standardized patients (i.e. trained assessors acting as patients) (Kurtz et al., 2017, pp. 272-273).

Research has shown that using real patients as assessors has several advantages, such as providing authenticity, which is a key aspect of contextual and experiential learning (Simek-Downing et al., 1986); however, when standardized patients (SP) are planned well, they can be indistinguishable from real patients, as one study shows (Rethans et al., 2007). Real patients also encourage students to do their best given that it is a real case; the case "*comes to life*" and becomes more "*enjoyable than a written case";* this motivates students to work harder (Bokken et al., 2009). Clinical education faces a significant challenge in integrating real patients into communication skills assessment. In a systematic review of communication skills teaching and assessment in dental training, the authors highlighted that (Carey et al., 2010, p. 70):

> *None of the studies which used real patients sought feedback from those patients, which is somewhat surprising given the possibility that the best people to inform students as to the acceptability of their communication style and hence validate assessment might be real patients themselves*.

Nonetheless, feedback provided by SPs can be more useful than that of real patients, as SPs tend to offer more detailed and constructive feedback in comparison to real patients who provide limited to no feedback (Bokken et al., 2009). Other advantages of SPs include their availability, flexibility and standardization (Barrows, 1993). Further, ethical concerns in involving real patients make SPs an acceptable alternative in certain circumstances.

2.4 SELF-ASSESSMENT

2.4.1. Conceptualization of self-assessment

There is a variety of terms in the literature used to refer to self-assessment in the context of health professions education, such as self-appraisal, self-evaluation, self-audit, self-rating and others, which, to an extent, all describe learner's engagement in assessing oneself (Duffy & Holmboe, 2006; Mays & Branch-Mays, 2016; Yates et al., 2022). Self-assessment was defined by Epstein et al. (2008) as "*a process of interpreting data about our own performance and comparing them to an explicit or implicit standard;*" however, this definition assumes that data are readily available for interpretation where in fact, data collection or selection are very critical subprocesses. Self-assessment in the lack of credible data is likely to be of little value (Duffy & Holmboe, 2006). Further, this

definition overlooks the importance of the context in which self-assessment occurs, the assessment method used, and, more importantly, the object of assessment.

A nuanced conceptualisation of self-assessment is to view it as a complex, multifaceted and multipurpose process in which the individual engages in generating/collecting and interpreting data about their performance in a particular domain guided by internal factors (biases, self-esteem, self-efficacy, goal orientation, self-awareness, self-monitoring, competence, internal standards etc.), inter-personal factors (relation to peers, instructors, and patients), contextual factors (learning environment, time, clinical climate etc.) and perceived understanding of some external standards (Andrade & Brown, 2016; Boud, 1999; Eva & Regehr, 2008; Sargeant et al., 2010). Some scholars conceptualised informed self-assessment as an "*externally informed reflective process*," meaning that external assessment and feedback given by supervisors, peers, self and patients are all part of the conceptual model (Sargeant et al., 2010, p. 1212). Nonetheless, the current study makes a distinction between self-assessment and external assessment to enable the investigation of each contribution before they become a tangled web (Sargeant et al., 2010).

Instead of going through every other possible definition of self-assessment, I will outline some important distinctions in the definitions used in the literature:

a) Global or analytic: Self-assessment can be global, where an overall judgement is made about performance, e.g., "*I am a good communicator*," or analytic, where a focused, narrow statement is made in a particular area, e.g., "*I listened attentively to the patient*"(Ward et al., 2002).
b) Ability or a pedagogical strategy: Self-assessment is either seen as a skill to be measured for accuracy or as an educational tool to enhance learning (Eva & Regehr, 2008; Yates et al., 2022).
c) Elicited or naturally occurring: Self-assessment can be elicited and formal, such as when a learner is tasked with assessing themselves using a structured form or natural/unsolicited (Sargeant et al., 2010; Yates et al., 2022).

Beyond these distinctions, self-assessment is also often confused with self-efficacy, which is confidence in completing a task (Bandura & Wessels, 1997).

2.4.2 Perspectives and debates in self-assessment

Debate in medical education has led to a decreased focus on self-assessment, largely due to evidence of a poor correlation between self-assessment and actual performance (Eva & Regehr, 2005; Eva & Regehr, 2008; Ward et al., 2002). Eva and Regehr (2008) argued that humans are predisposed to being poor at providing valid self-generated summative assessments of performance due to cognitive errors (e.g. Information neglect, memory biases, selective recall), sociobiological reasons (e.g. maintaining a positive outlook) and social reasons (e.g. not receiving feedback from others) (Bandura & Wessels, 1997; Dunning et al., 2004; Ende, 1983). In line with these arguments, it has been posited that looking outward can give more useful answers than looking inward (Eva & Regehr, 2008).

However, individuals' backgrounds, experiences, and beliefs shape their self-perception (Sargeant et al., 2010); because of that, the best approach might not be to "*convince them that external assessments are legitimate*" but to acknowledge their self-judgement, get in a similar mental space and have a dialogue with them that starts, acknowledges and respects their own self-assessment (Eva & Regehr, 2008, p. 18; Perera et al., 2010; Stalmeijer et al., 2010).

My readings of the research published in some esteemed journals in the medical education field, such as *Academic Medicine* and *Medical Education*, showed an interesting shift from the use of the term "self-assessment" favouring related like self-reflection and self-monitoring (Schaepkens et al., 2024; Shariff et al., 2020). This shift risks overshadowing self-assessment's value as a cornerstone of self-regulated learning (Sargeant et al., 2010; Zimmerman, 2000). Self-assessment, rightly understood, may not even need to be accurate " *to enable exercises in self-reflection to formatively facilitate performance improvement*" (Eva & Regehr, 2008, p. 18). The benefits of self-assessment may stem from the active involvement in the learning process, i.e. the accuracy of self-assessment is irrelevant as long as students engage in evidence-based reflection (Yan, 2018). However, accuracy in self-assessment is essential to enable meaningful revisions in practice. Self-assessment can also be a valuable resource for educators and learners; it could be a window to students' thinking, emotions, values, beliefs and perceptions of assessment criteria (Sargeant et al., 2010; Ward et al., 2002).

Moreover, self-assessment promotes democratic values by challenging traditional assessment power dynamics, fostering transparency, accountability, and collaboration between educators and learners (Cook-Sather & Matthews, 2021; Luong et al., 2024). It also supports the development of intrinsically motivated and self-regulated learners, as emphasized by educational theories like self-regulated learning (SRL) theory and self-determination theory (Williams et al., 1999; Yan, 2018; Zimmerman, 2000). The SRL framework, depicted in **Figure 3**, outlines a three-phase cyclical model: forethought, performance, and self-reflection (Zimmerman, 2000). In the forethought phase, learners set goals, develop strategies, and engage in self-efficacy. During the performance phase, they practice self-control by maintaining focus and engaging in self-observation. Finally, learners evaluate their work in the self-reflection phase, analyse outcomes, and self-react accordingly.

These phases are interconnected, meaning that changes in self-assessment can influence other processes, such as goal setting, motivation, and overall performance (Zimmerman, 2000). This highlights the critical role self-assessment is theorised to play in self-regulating learning.

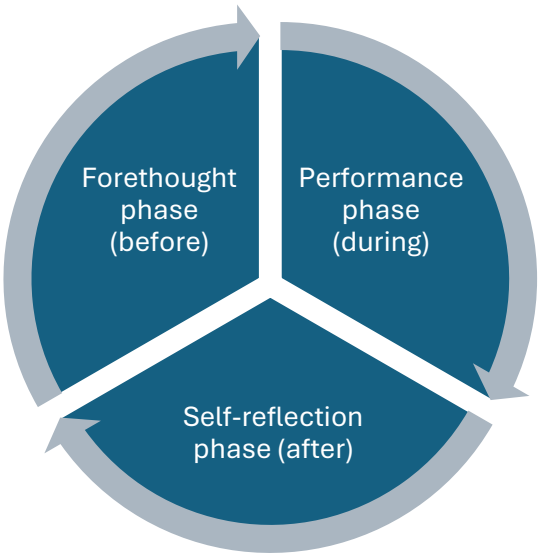


Figure 3. Self-regulated learning cyclical model adapted from Zimmerman (2000) and Artino (2012).

2.4.3 Self-assessment of communication skills

Self-assessment becomes particularly crucial in the development of communication skills. Recent research has shown that skilled communicators possess key attributes like self-awareness and reflective capacity, which are closely intertwined with the self-assessment process (Verheijden et al., 2023). This is also shown in the Calgary-Cambridge Guide for assessing communication skills, as one of the broad skills (perceptual skills) is about self-reflection and self-awareness (Kurtz et al., 2017). Self-assessment requires one to be aware of their norms, values, biases, and shortcomings in their communication. This mandates openness to feedback from the environment, be that the patient, nurse, or other sources; in this sense, one fosters a growth mindset, an intrinsic motivation to learn and experiment with various communication strategies (Verheijden et al., 2023), which might ultimately result in improved communication skills. Empirical evidence showed that self-assessment could play an effective role in communication skills learning (Perera et al., 2010). For example, Perera et al. (2010) developed an assessment approach for communication skills, which incorporates self-assessment and peer feedback with faculty and patient assessment in simulated settings and compared it to the traditional feedback model where only standardized patients and facilitators provide feedback. After one semester, the experimental group showed significantly higher communication skills scores (P=0.0001). The study concluded that self-assessment and peer feedback are effective educational strategies. However, the design did not distinguish the individual contributions of self-assessment and peer feedback, leaving it unclear whether one had more influence than the other.

There is an abundance of interventions to improve communication skills in health professions education literature, such as role-play activities, video-based techniques, cultural humility teaching, attitudinal learning, and simulation-based learning (Khalifah & Celenza, 2019; Kurtz et al., 2017, p. 107; Solchanyk et al., 2021). A notable critique of these methods is that they are primarily taught in classroom settings using lectures, visual aids, or standardised patients. However, within workplace-based learning environments where learning is deeply integrated with the context as per the socio-cultural theory (Hager, 2011; Mann, 2011), these predictable models of teaching and assessing communication skills become limited and might not always be transferable to real-world patient communication (Govaerts et al., 2013; Williams, 2022).

Learning within the clinical workplace is recognised as a dynamic, non-linear, and unpredictable process (Govaerts et al., 2013). From this perspective, self-regulated learning (SRL) becomes central to acquiring and maintaining clinical competence (White et al., 2013; Zimmerman, 2000). To engage effectively in this adaptive learning process, practitioners need to be reflective and engage effectively in revision and adaptation of performance. They need to recognise when to look up and when to slow down and make adjustments. Here, self-assessment combined with external feedback could become an effective learning approach (Perera et al., 2010).

Studies consistently indicate that dental students often struggle with self-assessing their competencies, especially in the area of communication skills (Alfakhry et al., 2022; Alfakhry, Mustafa, Ybrode, et al., 2023; Mays & Branch-Mays, 2016). Similarly, research in surgical training reveals that trainees tend to have less precise self-evaluation of their communication skills compared to technical skills (Moorthy et al., 2006). This poor correlation may be attributed to the challenges in assessing communication skills, which are inherently more personal and subjective compared to technical skills (Alfakhry, Mustafa, Ybrode, et al., 2023; Brown et al., 2017; Makoul, 2001b). Nonetheless, these results could also indicate a methodological flaw in operationalising communication skills. Many studies assess communication skills globally without specifying individual skills or tasks (Alfakhry, Mustafa, Ybrode, et al., 2023; Spoto-Cannons et al., 2019; Srikumaran et al., 2019). Global assessments can lead to biased judgments and cognitive errors (Eva & Regehr, 2011). Therefore, it is suggested that self-assessment should employ a detailed assessment instrument to better guide reflection, improve accuracy and increase educational impact (Sargeant et al., 2010; Yates et al., 2022).

These findings underscore the need for refined and validated assessment approaches to enhance self-assessment practices, particularly communication skills. That being said, the following section tackles the paradoxical topic of validity in self-assessment in educational assessment.

2.5 WHAT DOES VALIDITY MEAN IN THE CONTEXT OF SELF-ASSESSMENT OF COMMUNICATION SKILLS?

Validity of an assessment is defined as '*the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests*' (AERA/APA/NCME, 1999). This definition makes it clear that validity is not something

static a test has or does not, rather *"it is contingent on what a test is for, how it is used and how the results are interpreted" (Stobart, 2009, p. 162)*. This is different from previous conceptualizations of validity, which rely on the statistical property of a test; instead, it has now become an argument that starts and ends with the purposes of the assessment (Stobart, 2009). If multiple purposes exist, each purpose needs a validity argument and evaluation of its own. The outcome of a validity evaluation is not a binary determination (valid/invalid); instead, it sits on a continuum that reflects the degree to which the results fulfil their intended purpose (Stobart, 2009). Even a well-constructed test with robust statistical properties can be rendered invalid if results are misinterpreted or misused. Therefore, practical implications and social consequences have become central to the validity conversation (Linn, 1993).

There is an important distinction between the *assessment of communication skills* and *self-assessment of communication skills*. Even though they can overlap, these are two different constructs serving different purposes and are affected by different factors, one of which is the positionality of the assessor (Acheampong et al., 2023; Ward et al., 2002). Logically, the validity of the latter should not be hinged on its concordance with the former (Brown et al., 2015; Ward et al., 2002). Conclusions that self-assessment is poor, inaccurate, biased, and unreliable are founded on the purpose that self-assessment should mirror the formal assessment, where, in fact, it could be actually mirroring in part something else like the learner's own constructions and perceptions of their performance which in themselves are of pedagogical value (Sargeant et al., 2010).

3. MATERIALS & METHODS

3.1 CHAPTER GUIDE

This chapter is intended to:

- State the aims and research questions
- Present the overall mixed methods research design and link it to the aims
- Describe the assessment approach development
- Discuss and justify settings, sampling, data collection/generation and analysis within the quantitative and qualitative strands.
- Discuss the researcher's reflexivity and positionality
- Comment on rigour and ethical considerations

3.2 RESEARCH AIMS AND QUESTIONS

This study aims to develop and validate an assessment approach to self-assessment of communication skills in clinical settings in Syria. The assessment in this study aims to measure learners' *self-assessment of communication skills* to facilitate performance improvements and enable a meaningful dialogue with external assessment. This is my validity argument that I will try to establish through evidence. The following research questions were pursued to this end:

RQ1: How reliable is the developed assessment instrument in measuring self-assessment of communication skills?

RQ1a: How reliable are patients' assessments in comparison to self-assessments?

RQ2[1]: What is the assessment instrument's factor structure as reflected by the self-assessment scores?

RQ3: What is the impact of this assessment approach on facilitating learning and performance improvement?

RQ4: How do student participants' qualitative accounts explain their self-assessment scores?

[1] Assessing the factor structure as reflected by patient assessment was not feasible given the sample size of the patients pool as will be detailed later.

RQ4a: What factors explain alignment (or misalignment) of self-assessment scores with patients' assessment in the same clinical interaction?

3.3 MIXED METHODS RESEARCH DESIGN

This multi-phase research project utilised a mixed-methods (MM) design, combining quantitative and qualitative strands to answer the proposed research questions. The choice for MM stems from a pragmatist position that foregrounds *"the dictatorship of the research question"* and the suitability of methods in answering these research questions rather than linking the use of methods/methodologies with broader ontological and epistemological standings (Tashakkori & Teddlie, 2010, p. 17). While research questions one and two lend themselves to quantitative methods and data, research question three and four lend themselves to mixed methods, i.e., it can be investigated using both quantitative and qualitative methods/data. Mixed methods research was conceived to obtain a more complete and corroborated picture of self-assessment of communication skills as measured using the developed instrument (Creswell & Clark, 2017, pp. 7-11).

Sequential explanatory mixed methods research is the closest typology to this research design. Nonetheless, it does not fully capture it; therefore, I sought to describe it according to some common well-cited criteria: relative priority (weighting), timing, mixing procedures, and the purpose of the integration (Creswell & Clark, 2017; Greene et al., 1989). A summary of the specific MM design is provided in **Figure 4**, showing when and how quantitative and qualitative strands were integrated. While the first stage prioritised the quantitative element, I concurrently collected some qualitative data from participants during this stage. The second stage focused on qualitative interviews. Mixing happens in multiple ways, such as quantitative data informing participant selection or the interview questions. While the primary focus was on complementarity and expansion, other goals such as triangulation, development, and initiation were also pursued, albeit to a lesser extent (Greene et al., 1989).

The study began with instrument development, followed by pretesting for clarity and relevance in the Syrian context. Both qualitative and quantitative data were collected during assessment encounters. Quantitative analysis guided the selection of participants for the qualitative strand, with initial qualitative analysis of assessment comments also playing a role, though to a lesser extent. Assessment data informed the development of

the interview protocol, while qualitative analysis further refined the protocol and influenced certain quantitative data analysis decisions.

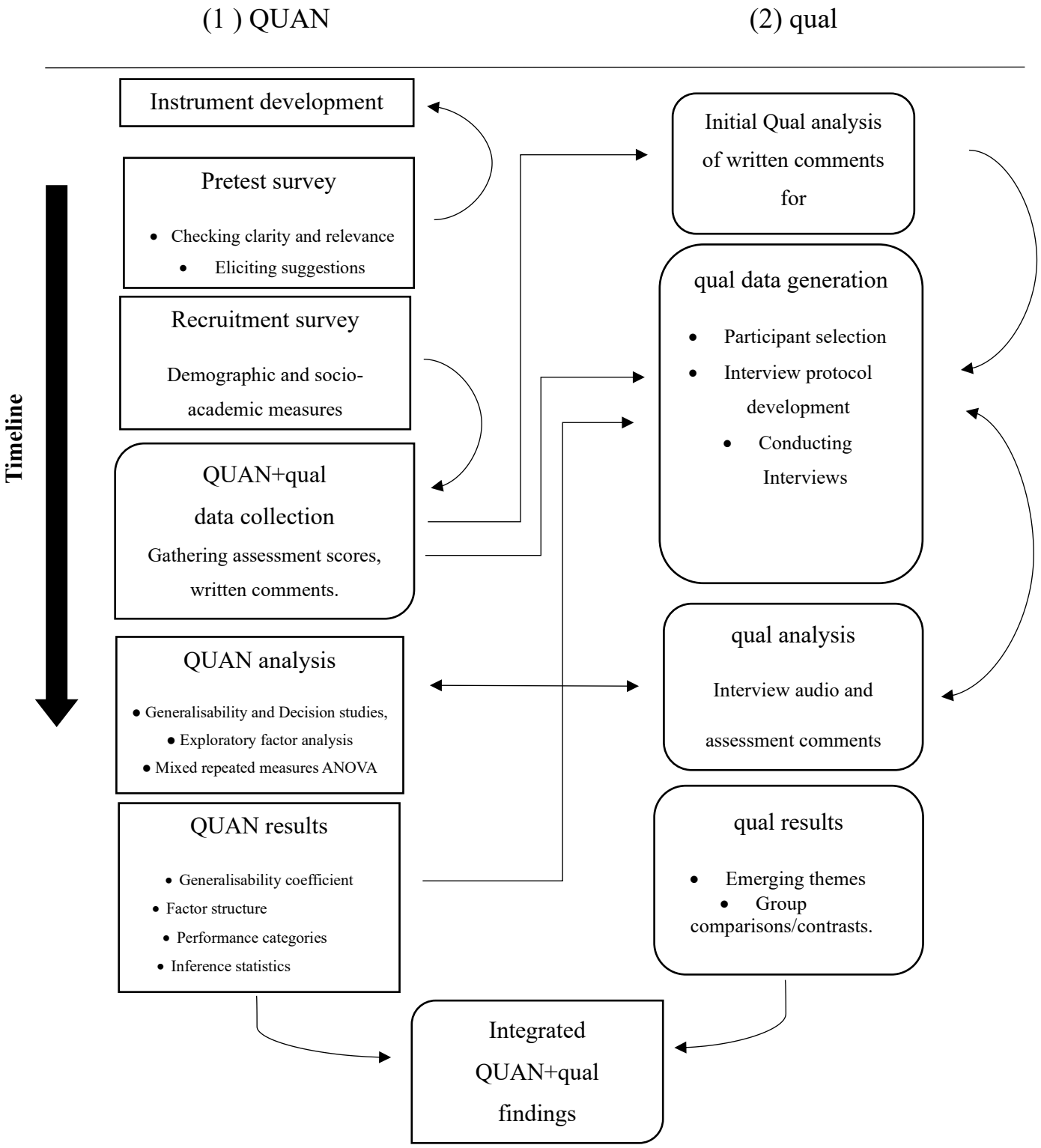


Figure 4. Diagram showing the specific research process and interactivity between QUAN and qual strands. A slight variation of Morse conventions was followed in the presentation (Morse, 1991). QUAN=quantitative; capital letters are used to show priority. qual=qualitative.

3.3.1 Remarks on methods integration

Integration can, as defined by Bazeley (2024, p. 7), as dynamic interdependence, evident in a connected, transactional, transformative, and coherent relationship between components from different methodological approaches to research.

Although sequential explanatory research typically separates quantitative and qualitative components, this study intentionally maximised the integration of both methods. The strands were interdependent, working together to achieve a common goal. Research question 3 employed a mixed-methods approach, combining both strands, while Research question 4 used qualitative data to expand on the quantitative findings. The strands were also interconnected, with the qualitative sample drawn from the larger quantitative sample based on analysis categories, which will be explored later in section 3.5. The quantitative assessment data informed the qualitative interviews and subsequent qualitative analysis influenced the analysis and interpretation of the quantitative data, which will be shown in sections 3.5 and 3.6.

3.4 ASSESSMENT APPROACH DEVELOPMENT

3.4.1 Assessment aim

The assessment aimed to capture how clinical dental students self-assess their communication skills in consultation sessions. Given the focus on communication and the fact that some patients are actors, no clinical intervention was done. Both patient and self-assessments were collected to compare perspectives and enhance the educational value.

3.4.2 Assessment instrument development

Medical literature on conceptualisation and assessment of communication skills was reviewed, including the E4 model, the Kalamazoo consensus, the SEGUE framework, the Communication and Interpersonal Skills (CIS) scale and the Calgary-Cambridge Guide (CCG) (Keller & Carroll, 1994; Kurtz et al., 2017; Makoul, 2001a, 2001b; Yudkowsky et al., 2006). Due to the overlap between frameworks, I selected the most comprehensive framework, popularly used and well-described, which was the Calgary-Cambridge Guide (CCG) (Kurtz et al., 2017; Kurtz et al., 2003; Kurtz & Silverman, 1996; Silverman et al., 2016). The CCG was originally developed in 1998 and later revised in 2003 (Kurtz et al., 1998; 2017, pp. 39-40).

The CCG guide, originally consisting of 71 tasks organised by Kurtz et al. (2017), was streamlined to a manageable set to ensure the form can be completed within a reasonable timeframe. I began by summarising each section and selecting the most relevant tasks to achieve the stated objectives of each CCG communication task (**Figure 2**). Similar tasks were merged, such as combining "greets the patient" and "introduces self" into "Greets you and introduces self appropriately." Tasks with overlapping objectives, such as "*accepts the legitimacy of patients' views*," "*uses empathy*," and "provides support," were consolidated into "*Demonstrates genuine care and responds with empathy*" (Kurtz et al., 2017, pp. 44-48). This approach reduced the list from 71 to 26 items.

Two additional items, "*Acknowledging deficiency in knowledge or skill*" and "*Providing safety nets*", were added after consultation with a supervisor with a medical background. The former item, not originally in the CCG inventory, was included because dental students may encounter cases beyond their expertise, making it crucial for them to recognise and address this from a self-assessment perspective while maintaining patient confidence.

The original section, "*Gathering Information,* " was renamed "*Exploring the Problem with the Patient,* " emphasising a patient-centred approach. The sections "*Providing Structure*" and "*Building the Relationship*" were merged into a new section called "*Cross-stage items*" (**Figure 5**). Finally, my dissertation supervisors and I reviewed the items, leading to further refinement, ensuring clarity, conciseness, consistency and suitability for measurement.

Since the CCG was not originally developed for dental practitioners, I reviewed dental education literature to ensure alignment and relevance to dental practice. After careful review, all items were confirmed to be relevant, as they are referenced in various dental academic sources (Khalifah & Celenza, 2019; Salim et al., 2023; Schönwetter et al., 2012). It is also worth noting that a variation of CCG has been used previously to assess communication skills in dentistry (Salim et al., 2023). The first version of the instrument developed was in English (**Appendix I**), necessitating appropriate cross-cultural validation into Arabic, which is explored in the following section.

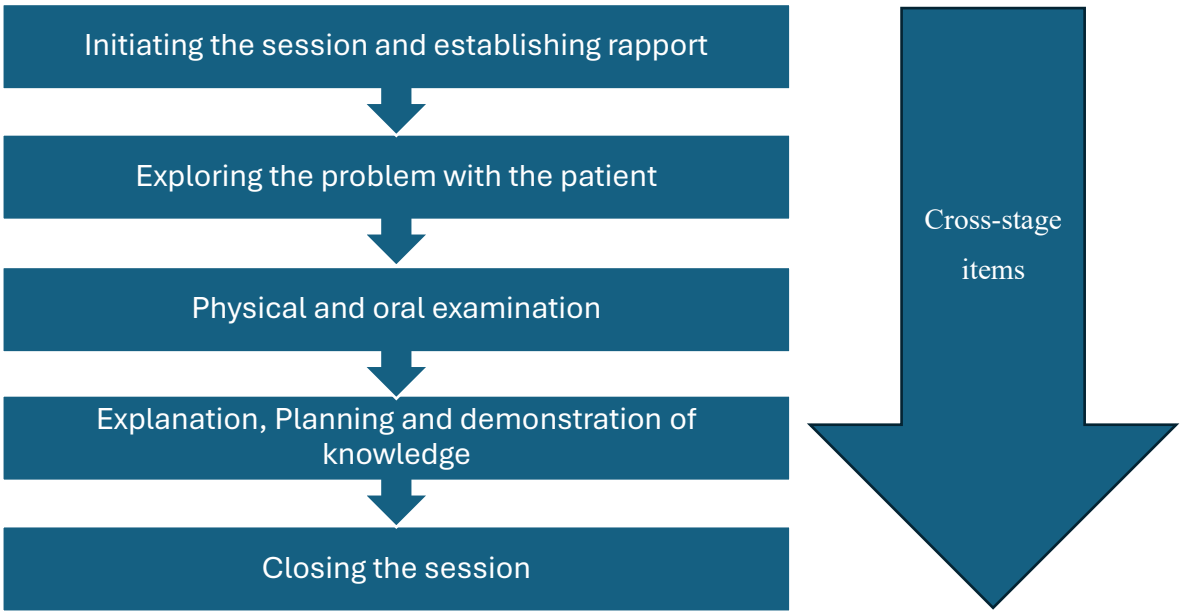


Figure 5.Diagram showing the adapted framework from Calgary-Cambridge Guide used in this study.

3.4.3 Translation and cross-cultural validation into Arabic

The assessment instrument was to be employed in Syria, where the mother tongue is Syrian Arabic. Therefore, the instrument was translated and cross-culturally validated into Syrian Arabic following the well-accepted Sousa and Rojjanasrirat (2011) 7-step guideline to overcome the language barriers and ensure cultural relevance. First, three independent forward translations from the source language (SL) to the target language (TL), which in this study were conducted by three people: Myself, a specialist physician with expertise in formal TL (Fuṣḥā /standard written Arabic), and a practising dentist with knowledge of colloquial TL. Second, translated versions were compared and discussed amongst the translators to resolve any discrepancies or ambiguities. Third, blind back-translation was conducted by a different independent translator with medical training and knowledge of TL and SL. Fourth, a comparison of the back-translated and original SL was made, and revisions were done. Fifth, the pre-final version instrument was pilot-tested on a monolingual sample, detailed in Section 3.4.4 below. Lastly, the instrument was applied to a sample of dental students, and full psychometric testing was conducted (described in section 3.5.4). Preliminary psychometric testing, which should be done before full psychometric testing, was skipped due to the lack of accessibility

to a bilingual population. According to Sousa and Rojjanasrirat (2011), Step 6 is rarely used.

3.4.4 Pilot testing of the translated instrument: Clarity and relevance

The pre-final translated version (P-FTL) of the instrument was pilot-tested with 37 participants, including clinical dental students, patients, and experts at Damascus University. A sample size of 10-40 was recommended by previous literature, and this study achieved 37 participants (18 patients, 17 clinical dental students, and two experts) (Beaton et al., 2000; Sousa et al., 2009). A specialist and a lecturer (experts) were asked for feedback as they were regarded highly in the university for their skilled communication and professional attitude. Feedback on clarity and relevance was gathered through paper-based questionnaires. Most items met the 80% clarity and relevance thresholds (Sousa & Rojjanasrirat, 2011), except Item 14, which was retained due to its near-threshold score and expert support. Relevance is conventionally tested by a panel of experts to measure the Content validity index (Shi et al., 2012), was evaluated by patients, clinical dental students, and experts. This inclusive approach was adopted because, from a healthcare and educational perspective, it is crucial to understand patients' and dental students' perceptions of relevance (Moult et al., 2021; Sambell, 2013). Participants' characteristics and detailed results are in **Appendix II**. Based on expert feedback, two additional items were added: "*providing preventive and treatment instructions"* and "*Encouraging the patient to ask questions about anything."*

3.4.5 Final assessment instrument

The final assessment instrument consisted of 30 items and one overall judgment item, organised into six sections: the five examination stages and a cross-stage items section adapted from the CCG guide (**Figure 4**). Each section included space for comments to clarify ratings and provide qualitative feedback and an additional area for comments on communication skills not covered in the form to ensure comprehensiveness (content validity). The Arabic and English assessment forms are provided in **Appendix III**.

Additionally, self- and patient assessors were asked to identify one area of improvement and one area of excellence in four main areas: patient-centred care, effective communication, respect and empathy, and professionalism and organisation. This section, which was included for its educational value and was not analysed in the study.

Standardised patients had two extra yes/no questions about the accuracy of diagnosis and treatment, which was not incorporated into the analysis in this study. Two versions of the assessment form were created: one for self-assessing students and one for patient assessors, differing only in pronoun use (first-person for students, third-person for patients).

A 9-point scale with three anchors (Needs improvement, Acceptable, Excellent) was used. While some assessment forms in medical education literature used “Fail” or “unsatisfactory” as an anchor, I opted to use “needs improvement” to avoid any negative morale effects on students and to encourage a focus on future-oriented actions (Yon, 2020). A 9-point scale was used instead of a 5-point scale to provide greater granularity and sensitivity, allowing for more precise differentiation in performance levels. This choice aligns with evidence that longer scales offer higher validity, better discriminating power, and convey more useful information (Cook & Beckman, 2009; Preston & Colman, 2000). Nonetheless, there are also scholars who put forward arguments for having fewer scale points (Shrout & Yager, 1989).

3.4.6 Assessment approach: procedure and patient assessors

The assessment involved six consultations with six different patients: three standardised patients (SPs) and three real patients (RPs). The SPs, all qualified dentists, were trained in using the assessment instrument and practised on two dentists, resolving any ambiguities. Both SPs and RPs were diagnosed by two general dentists and a Year 2 Oral Medicine specialist. SPs were informed of their diagnoses before assessing students, while RPs remained unaware of their diagnoses until the experiment's conclusion to maintain realism. SPs concealed their identities as dentists, and students were only told that some patients were SPs and dentists, not their identities.

Diagnostic information was not collected as research data but was used to ensure case suitability. Panoramic radiographs were taken for all patients and made available to the students. Two of the SPs were male, and all real patients were female. One RP was an undergraduate, another postgraduate, and the third was a housewife with a pre-high school education. The other two RPs were student health professionals but not in dentistry.

The use of six cases/raters was based on previous research indicating that five encounters are needed for sufficient reliability in a mini-clinical evaluation exercise (Wilkinson et al.,

2008). On assessment days, all patients were seated in a single room with six dental units/chairs. The assessment occurred from 9:30 AM to 12:30 PM, with six students rotating among the patients. Each consultation was 20 minutes, followed by 10 minutes to complete the assessment forms. After completing all consultations, students reviewed anonymised patient assessments and feedback.

Required materials included gloves, surgical masks, and an examination toolkit (probes, tweezers, dental mirrors). All materials, radiographs, and research costs were covered by me.

3.5 QUANTITATIVE STRAND: SETTINGS, SAMPLE, AND METHODS

3.5.1 Setting and student participants

This study was undertaken at the Department of Oral Medicine, Faculty of Dental Medicine, Damascus University (DU) in Damascus, Syria.

To calculate the sample size, I used G*Power software with the following configurations: ANOVA repeated measures as the statistical test, medium effect size (Effect size F: 0.25), an alpha of 0.05 and a power of 95% (**Appendix IV**). The effect size has been approximated using previous self-assessment studies conducted at DU dental school (Alfakhry et al., 2022; Alfakhry, Mustafa, Ybrode, et al., 2023). The ANOVA test was chosen as it is the basic test for the Generalisability study, the main focus of this study (Brennan, 2001). A calculated sample size of 54 was given by the program but to account for any dropouts I aimed to recruit 60 student participants. The inclusion criteria for student dentists were: being a 5th-year dental student upon completion of the recruitment survey, being available during the selected assessment day and consenting to participate in the study.

The student participants were approached using the river sampling technique, where random sampling would not have been feasible due to access constraints (e.g., contact information and full sampling frame). River sampling is an online sampling technique where websites such as social media platforms are used to recruit participants (Lehdonvirta et al., 2021). I posted to recruit participants on the official 5th-year dental students' Facebook page. A total of 82 dental students completed the recruitment survey out of approximately 700 possible students; I contacted 74 students who were available

during the set assessment days; 19 participants apologised due to changes in their schedule or being busy. Ultimately, 55 students participated, of which one was later excluded from analysis as they only completed 5 out of 6 patient encounters.

Table 1 (self-report data) shows participants' characteristics. Sex was dichotomous (male/female) based on the participant's biological sexual self-identification. These choices were deemed appropriate due to the cultural context.

Table 1. Student participants' characteristics (n=54)

Sex*	54% Female				
Age	22.5±0.5				
Family income	13% (Below average)	39% (Average)	(43%) Good	(6%) Very good	
University grade	54% (75-80 band)	28% (80-85 band)	13% (85> band)		
Social self-perception	19% (introvert)	80% (extroverts)			
Social activity per week	(0%) Never	(11%) Rarely	(29.6%) Sometimes	(29.6%) Often	(29.6%) Frequently
Communication skills	(4%) Needs improvement (1-3 score)	(50%) Acceptable (4-6 score)	(46%) Excellent (7-9 score)		

*sex was binary due to the cultural context.

3.5.2 Data collection

Assessment encounters were carried out during 10 working days between April 28 to May 22, 2024 during 10 working days. I was present during three assessment days and had to relay assessment supervision duty for the remaining 7 days to a trained research assistant. Due to scholarship travel restrictions, I could not stay in Damascus to supervise data collection in person as originally planned.

Each student completed six encounters, during which both self-assessment and patient assessment were conducted. Students were intended to complete all 6 patient encounters back-to-back during one assessment day. However, due to the absence of one of the patients on two days, 12 students were told to come on two different days to ensure completion of 6 patient encounters. During the assessments, a supervisor (me and later a research assistant) was present to guide students through the process, prepare examination kits, call turns, and ensure that students and patients adhere to the time limit and research

protocol. The number of assessment encounters carried out was 324, and since this involved paired self- and patient assessment, there were 648 completed assessment forms.

All data was processed and entered into Microsoft Excel 365 (Version [2407]; Microsoft Corporation, 2023) and later exported to R (Version 4.4.0; R Core Team, 2024), RStudio (RStudio Team, 2024) and IBM SPSS Statistics (Version 29; IBM Corp., 2023) for analysis.

3.5.3 Missing data and outliers

One student's data was removed from the dataset due to incomplete responses, leaving a balanced dataset of 6 self- and patient assessments per student.

Item non-response: Self-assessments had minimal missing data, ranging from 0.3% to 3.4% per item, with item 20 showing 5.6% missing responses. Patient assessments had even fewer missing responses. Missing data analysis is shown in **Appendix V.** Overall, missing data points account for 0.9% of the total.

Due to the low percentage of missing data (<1%), median imputation was used, as it better maintains central tendency in slightly right-skewed data (Schumacker, 2014). Imputation of missing data was necessary to perform a balanced design for the reliability analysis (Andersen et al., 2021)

Outliers were identified using z-scores, with an absolute z-score greater than 3 flagged as an outlier (Barnett & Lewis, 1994). The number of outliers was minimal, with 0 to 6 per variable in self-assessments and 0 to 10 per variable in patient assessments. Deleting outliers had a minor effect on the mean and standard deviation, so they were retained for the analysis, as they represent true variations and not errors (**Appendix V**) (Aguinis et al., 2013; Barnett & Lewis, 1994; Zijlstra et al., 2011).

3.5.4 Data Analysis

Generalisability study followed by a Decision study

Reliability is traditionally measured using Classical Test Theory (CTT), a psychometric framework for evaluating measurement instruments like tests and surveys (Bloch & Norman, 2012). In contrast, Generalizability Theory (G Theory) considers multiple

sources of variance, called facets of generalisation, with the object of measurement (usually the person or p) as the facet of differentiation (Brennan, 2001).

This study examined two main facets of generalisation: a random facet (cases or raters) and a fixed facet (items). The design is balanced and fully crossed: students (p) × items (i) × cases (c) for self-assessments and students (p) × items (i) × raters (r) for patient assessments. In both, the student (p) is the facet of differentiation (**Figure 6**). Fully crossed refers to a study design in which every level of one facet is paired with every level of the other facets.

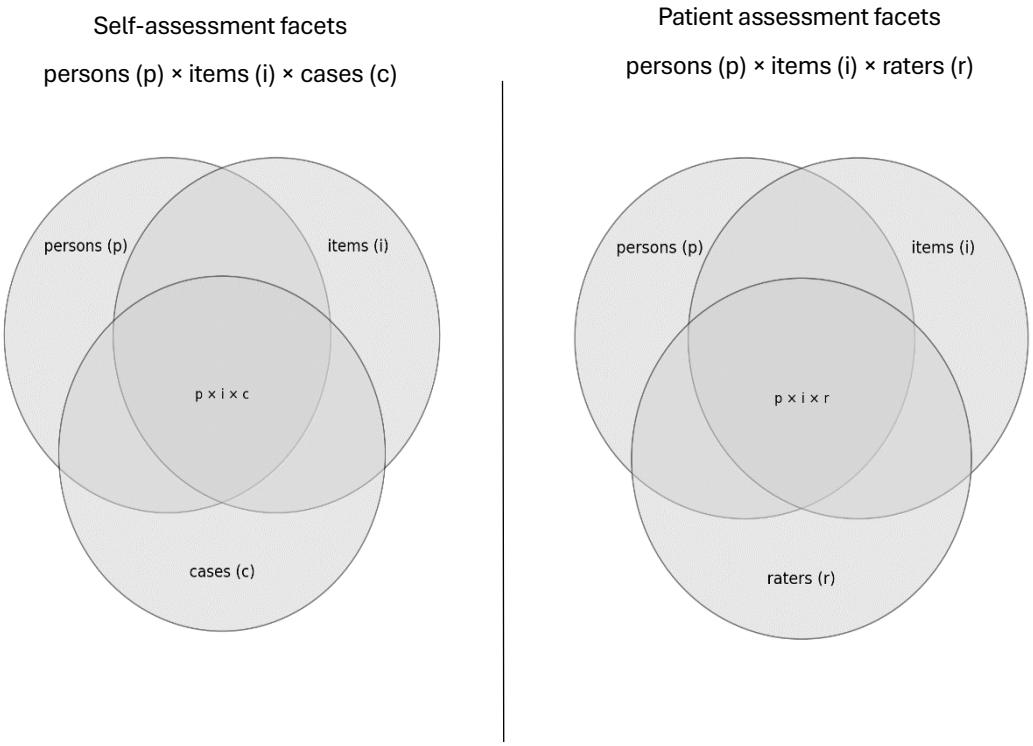


Figure 6 Two Venn diagrams for self-assessment (right) and patient assessment (left) showing a fully crossed two-facet design: (p × i × c) and (p × i × r).

The G coefficient, ranging from 0 to 1, is similar to the Intraclass Correlation Coefficient (ICC). A G coefficient of 0 indicates no generalizability across facets, with most measurements due to error, while a value of 1 indicates perfect generalizability. For summative exams, a G coefficient of 0.8 or higher is considered acceptable; for formative assessments, 0.6-0.7 is typical (Bloch & Norman, 2012; Wilkinson et al., 2008). A caveat in interpreting the G-coefficient for the self-assessments is that the rater effect is conflated with the student (p) effect, making the interpretation of a G-coefficient vague; therefore,

the G-coefficient for the self-assessment must be interpreted cautiously. Nonetheless, it is useful to investigate the variance attributed to cases and items.

All 31 items (30 analytic + 1 global) were included in the G study, as including the global item has been shown to increase reliability (Daniels & Harley, 2017; Hodges et al., 1996). The second component of G Theory, the Decision (D) study, identifies optimal conditions for higher reliability.

The G and D studies were conducted in R (Version 4.4.0) and RStudio using the gtheory package (Moore, 2016) following Huebner and Lucht (2019) guide. **Appendix VI** provides an example code for performing G and D studies for self-assessment data (p × i × c).

Exploratory Factor Analysis

To address research question two, Exploratory Factor Analysis (EFA) was used to evaluate the structural validity of the assessment instrument and identify underlying constructs (factors) that explain the correlations among the items based on self-assessment data (Watkins, 2018)

While factor analysis typically requires a large sample, the adequacy of sample size depends on factors such as the strength of correlations, factor overdetermination, the number of variables, and communalities (Field, 2017, p. 797; Velicer & Fava, 1998). These factors can interact, allowing one weakness to be offset by another strength (Velicer & Fava, 1998, p. 248).

Although our sample size was small (n=54), I proceeded with factor analysis, assessing the adequacy of the sample size using several measures. The Kaiser-Meyer-Olkin (KMO) measure was above 0.8 for each measurement (patient) and 0.961 for the entire dataset, indicating "marvellous" sampling adequacy (Kaiser & Rice, 1974). Bartlett's sphericity test was significant ($P < 0.001$), suggesting that the correlations between variables (items) overall are significantly larger than zero (Field, 2017, pp. 798-799). Communalities were mainly in the 0.6-0.7 range, suggesting that the sample size might be adequate (Field, 2017, pp. 788, 797). Communalities are shown in the results section. Additionally, factors with four or more loadings greater than 0.6 are considered reliable, regardless of sample size (Guadagnoli & Velicer, 1988). Based on my judgement of these elements, I determined that the factor analysis could be appropriate.

A limitation of applying EFA in this study is its inability to deal with the multi-level structure of the repeated measures of self-assessment. To mitigate this limitation and support EFA's reliability, I split the data in half and conducted EFA for each half.

Principal Axis Factoring was chosen as the extraction method due to its suitability for smaller samples (de Winter & Dodou, 2012). Direct oblimin rotation, a non-orthogonal method, was used because Promax is more suited to very large datasets (Field, 2017, p. 794). The number of factors was determined using multiple criteria, including the scree plot, eigenvalues, additional explained variance, and the number of variables per factor (Schönrock-Adema et al., 2009). More specific thresholds for these criteria will be mentioned in the results section.

Further quantitative analysis

I conducted additional analysis to explore patterns in self-assessment and patient assessments. A mixed repeated measures ANOVA was used to assess the possible actual impact of the assessment method considering the effect of patient order and student participant characteristics (gender, family income, academic performance) on both self-assessments and patient assessments. A *P*-value of ≤ 0.05 was considered significant.

Students were ranked based on the aggregated six patients' scores (the sum of scores across all 31 items) and then divided into three equal tiers (low, mid, and high) for comparison. The same ranking process was applied to self-assessment data, though these ranks were only used for sample selection in the qualitative phase (details in the next section). **Table 2** shows the aggregated score ranges for the tiers.

I initially planned to use absolute and arithmetic differences to measure self-assessment accuracy, as per Fitzgerald et al. (2003). However, qualitative interviews revealed that patients and students assessed performance from different perspectives and used the assessment instrument differently. For example, a score of 7 from a student may not have the same meaning as a 7 from a patient, so a difference of zero could misleadingly suggest agreement. Therefore, I decided to forego these measures and relied on correlation analysis instead.

Table 2. Score ranges and number of participants based on their tier on self-assessment and patient assessment data.

	Patient assessment		Self-assessment	
Group	N.	Score range	N.	Score range
High tier	18	1413-1459	19	1410-1645
Mid-tier	19	1341-1408	18	1225-1405
Low tier	17	1129-1340	17	903-1222

3.6 QUALITATIVE STRAND: SETTINGS, SAMPLE, AND METHODS

3.6.1 Setting and participants

Student participants for the qualitative strand were a purposive subsample of those who took part in the assessment (n=54) and were selected based on criteria derived from the quantitative analysis. After categorising participants into high, mid, and low tiers based on self-assessment (SA) and patient assessment (PA) data, four groups were created:

1. Low PA, low SA group
2. low PA, high SA group
3. high PA, high SA group
4. high PA, low SA group

Table 3 shows the distribution of self-assessment high, mid, and low groups across their corresponding patient assessment groups. Since the fourth group had only one participant, the nearest misaligned participants from the mid-SA group were included. Further selection within the low-low and high-high groups, each containing 10 participants, was refined using written comments from self-assessments and patient assessments to identify the most relevant cases, such as those with detailed reflections or unique feedback. The mid-tier group was excluded to focus on extreme cases, which are more likely to reveal critical insights and are more relevant to the research question of alignment/misalignment. This approach also maximised the efficient use of time and resources within the constraints of an MSc research project.

All invited participants accepted the interview invitation except for one, who was replaced with another from the same pool. In one instance, a participant agreed only to a text-based interview; however, due to the limited information provided, another participant from the same pool was recruited, and the text-based case was excluded from

the analysis. Due to unforeseen technical issues (unnoticed recording malfunction at the beginning), one interview was regrettably unrecorded. Nonetheless, the notes taken during this interview were used to develop the interview protocol; they were not used in the analysis due to their brevity. Another interview was conducted with someone from the same group to replace the unrecorded one.

Participants were unaware of the specific selection criteria, and those from the low PA group were unaware of their ranking. All interviews, conducted via Microsoft Teams (Microsoft Corporation, 2024) between June 21 and July 4, 2024, lasted 40-60 minutes, with the interviewer (me) in Oxford, UK, and the participants in Syria.

Table 3. Distribution of self-assessment high, mid, and low groups across the three counterpart patient tiers.

PA tier	SA tier	*N.*	*%*	*N. of interviewees*
Low (n=17)	High	3	18%	3
	Mid	4	24%	-
	Low	10	59%	3
Mid (n=19)	High	6	32%	-
	Mid	7	37%	-
	Low	6	32%	-
High (n=18)	High	10	59%	3
	Mid	7	41%	2
	Low	1	6%	1

PA=patient assessment, SA=self-assessment.

3.6.2 Data generation and analysis

A preliminary interview protocol was developed in English, focusing on understanding how students self-assess, the perceived possible impact of self-assessment, reasons for alignment or misalignment with patients' assessments, and factors influencing self-assessment. Since the interviews were conducted in Syrian Arabic, the protocol was translated and double-checked for accuracy in the Syrian dialect. The English version, along with any refinements during the interviewing process, is provided in **Appendix VII**.

These semi-structured interviews were tailored to the participants' self- and patient-assessment results. Before each interview, participants' data, including SA and PA scores, aggregated scores, and written comments, were prepared in an Excel sheet. Key patterns, such as consistency or discrepancies in SA and PA, were highlighted for discussion. Initially, participants received individual pseudonymised PA data, but this was changed to only average PA scores to protect patient identities after it became apparent that interviewees could guess the patients based on their scores.

Participants were asked to be in a private space during the interview; all complied except one, who chose a café. Interviews were audio-recorded, as technical and consent issues limited video recordings. Audio recordings were imported into NVivo 14 (QSR International, 2024) and coded without full transcription due to time constraints. Selective transcription was used when necessary, as transcription software struggled with non-standardized Arabic dialects (El Gayar & Suen, 2018). Direct coding was supported by prior studies showing evidence of reliability (Gravois et al., 1992; Mayerovitch, 1972)

A variation of Braun and Clarke (2006) version of thematic analysis was used for its flexibility. The steps included familiarising the data, coding, constructing themes, and reviewing and defining themes. Top-level categories were based on the research questions, while emergent themes arose from participant responses. During the analysis process, research question four was tweaked a little as the original question was limited and was not the best version to capture what was being told by the interviewees (Braun & Clarke, 2006; Terry et al., 2017). The original RQ4 was: How do student participants explain the alignment (or misalignment) of their self-assessments with patient assessments of the same clinical interactions? This was changed to the current RQ4 and RQ4a, shown in section 3.4.

The analysis began inductively, with initial themes emerging from the data. Two themes were later developed deductively using Self-Regulated Learning (SRL) theory. Pseudonyms, chosen as random Syrian Arabic names with appropriate gender, were assigned to all interviewees to protect their identities and respect cultural norms. The qualitative findings and the used anonymized quotations were shared with interviewees for them to comment or make any modifications to their statement if they wish; interviewees did not make any changes to the final report.

3.7 REFLEXIVITY AND POSITIONALITY

3.7.1 Theoretical stance

At the start of this chapter, I mentioned that this research design is rooted in a pragmatist stance. However, it aligns more with methodological pragmatism than with an epistemological one, as distinguished by Maxcy (2003). I do not identify as a post-positivist, believing in a single reality measurable through statistics, nor as an interpretivist, believing in multiple socially constructed realities (Lincoln et al., 2011, pp. 114-115). choose not to engage deeply in the "paradigm wars" or firmly position myself epistemologically.

This research on self-assessment of communication skills challenged my ontological and epistemological views. While the assessment aims for objectivity, students interpreted and used the scale differently. However, statistical methods revealed consistent patterns despite these varied understandings. In the qualitative strand, where participants' self-assessment scores conflicted with their verbal explanations, I realised that both words and numbers can be deceptive. Even standardised patients, trained to assess, were influenced by their values and backgrounds, especially in personal domains like communication skills.

My evolving theoretical position is that there is objectivity in the subjective and subjectivity in the objective. Only through critical engagement can one discern what is meaningful and useful. This view aligns with Pring (2000) critique of "false-dualism". I support a pluralistic approach, where multiple realities and perspectives coexist (Johnson & Onwuegbuzie, 2004). Within this framework, knowledge is valuable if it is useful and practical, serving to solve problems and achieve outcomes. Truth is defined by what works in practice, consistent with the pragmatism paradigm, which is succinctly reflected by what John Dewey succinctly wrote that "*in order to discover the meaning of the idea, [we must] ask for its consequences*" (Dewey, (1948, 1920), p. 132).

3.7.2 Personal and professional background

I graduated from Damascus University with a degree in dental medicine in 2020 after six years of study. Following graduation, I took on various roles at the school, including research and supervision positions. My strong relationships with faculty and supervisors greatly facilitated my access to the Department of Oral Medicine and 5th-year students

for this research. Additionally, my identity as a researcher from Oxford University helped legitimise the study's credibility among stakeholders and participants.

Being a former student at the school allowed me to make meaningful interpretations of the research findings, some of which I anticipated, such as the low priority given to communication skills and the dominance of the dentist in patient interactions. However, recognising these assumptions helped me identify exceptions to them. My relationship with participants was nuanced, balancing an insider-outsider perspective. My insider status often helped build trust, though in some interviews, I was treated as an outsider, especially when participants directly asked if I was in Syria or had studied at the same dental school.

3.8 COMMENTS ON RIGOUR

For the quantitative strand, a traditional validity framework was adopted (De Vaus, 2001; Frambach et al., 2013):

1. Internal validity: Ensured through robust sample size calculation, detailed contextualisation of the educational environment, and standardised assessment conditions.
2. External validity: Supported by repeated measures, varying assessment conditions, ecological validity of the clinical context, and verification of structural validity.
3. Reliability: Assessed using a generalizability study, estimating various sources of measurement variance.
4. Objectivity: Enhanced by anonymizing participants during assessment, analysis, and reporting. The dataset was safeguarded and made openly available, ensuring transparency and allowing verification and reproduction of findings.

While some validity criteria overlap with the assessment approach's validity, it is important to distinguish between them. The assessment approach also includes unique criteria, such as educational impact and usefulness.

For the qualitative strand, I followed Lincoln et al. (2011, p. 122) conceptualisation of rigour: credibility, transferability, dependability, and confirmability. Practices that enhance trustworthiness, such as transparency and reflexivity, were applied across both

qualitative and quantitative strands. Transparency, achieved through honest and deep reflexivity, supports both credibility and confirmability in qualitative research and objectivity in quantitative research. Objectivity, as defined by (Bird, 2020), involves honesty in representing findings, using measures and references that allow for comparison and verification, and practising reasonable judgment. I also employed data triangulation (assessments, written comments, interviews), member checking, and supervisor feedback to strengthen our findings.

For the mixed-methods component, legitimation was established through sample integration, multiple validities, and weakness minimisation (Onwuegbuzie & Johnson, 2006). Sample integration was achieved by recruiting qualitative interviewees from the quantitative sample, supporting the quality of meta-inferences. Multiple validities legitimation involved establishing the relevant validities for both quantitative and qualitative strands and the mixed-methods approach. Weakness minimisation was accomplished by offsetting the weaknesses of one strand (e.g., lack of understanding of student score interpretations in the quantitative strand) with the strengths of the other (e.g., in-depth exploration in the qualitative strand). Similarly, the lack of generalizability in the qualitative strand was balanced by the strengths of the quantitative strand.

3.9 ETHICAL CONSIDERATIONS

This research and its three amendments were approved by the Education Departmental Research Ethics Committee (DREC) at the University of Oxford (REF: EDUC_C1A_24_096) (**Appendix VIII**). Although Damascus University lacks an ethical review board for educational research, the head of the Department of Oral Medicine reviewed and approved the research plan[2]. As no NHS patients, staff, faculty, or data were involved, NHS Research Ethics Committee approval was not required. A CUREC 1A application was prepared and approved by DREC, and three subsequent amendments were also approved due to changes in the research design. For a summary of the amendments, please see **Appendix VIII**. Consent forms and participant information sheets in English are provided in **Appendix VIII**. This project adhered to the British

[2] For Correspondence: Prof Omar Hamadah, Omar.hamadah@damascusuniversity.edu.sy

Educational Research Association (BERA) guidelines, respecting participants' autonomy, privacy, and other rights (BERA, 2018).

Informed consent

There are two main activities that constitute informed consent: "*participants need first to comprehend and second to agree voluntarily to the nature of their research and their role within it*" (Israel & Hay, 2006, p. 61). Consent was obtained from three participant groups: pilot testers, patient assessors, and students. Each received an information sheet in Arabic outlining their role, risks, benefits and a consent form to sign (**Appendix VIII**). It was emphasised that participation was voluntary and that they could withdraw at any time without explanation.

Confidentiality

Complete confidentiality was maintained for all participants. Identifying information was redacted, and randomised codes were used for assessments. The master list was securely stored on Microsoft 365. Students were given access to pseudonymised patient feedback for educational purposes, with patients instructed to provide professional and constructive comments. During interviews, only average patient assessment scores were shared to protect patient identities. Two students who received anonymised individual patient scores were able to somewhat guess patient identities based on scores, though they were not absolutely sure. Aggregated patient averages were used afterwards to protect patient identities.

Students were informed of the presence of standardised and real patients but were not given their identities. This was to maintain authenticity in student behaviour during interactions, as opposed to an incognito patient assessor approach, which involves deception and would require a waiver of informed consent (Rhodes & Miller, 2012).

Avoiding harm and maximising good

While in bioethics, research harm is often thought of as "*physical, psychological, social, or economical damage"*, in social sciences it is often "*psychological distress, discomfort, social disadvantage, invasion of privacy or infringement of rights*" (Israel & Hay, 2006, p. 96). Although this research was considered low risk, certain potential discomforts or harms were mitigated throughout the process.

Since the research took place in clinical settings involving patient examinations, the highest standards of infection control were maintained, with a supervisor (me or a research assistant) always present. To prevent patient fatigue, examinations were spread over 10 days, limiting each patient to no more than six examinations per day (three hours total). This also helped prevent assessor fatigue (Paravattil & Wilby, 2019). Patients were compensated for their time, receiving 4,000 Syrian pounds per student, totalling 240,000 Syrian pounds for 60 student participants, even though only 54 completed the assessment. Patients, student participants and interviewees all received a certificate of appreciation as a token of my gratitude for their time and work.

The research included an educational component that many students found valuable, reporting new skills learned, as reflected in the results. Reflection on communication skills and patient feedback positively impacted their learning. Students received an electronic copy of the self-assessment form for future use and had direct access to me and the supervisor for emotional support if needed. During interviews, care was taken to avoid discomfort, especially for those who did not perform well in patient assessments. To mitigate this, discussions focused on future-oriented feedback, balancing strengths with areas for improvement. This approach was effective, as interviewees did not express discomfort based on their interactions and feedback. Furthermore, the final report was shared with interviewees who were allowed to comment or modify their quoted statements if they wish it- none the of the interviewees requested any change in the report.

3.10 ACCESS TO RESEARCH DATA AND MATERIALS

The de-identified quantitative dataset is available from the author, Ghaith Alfakhry (ghaithalfakhry@gmail.com), upon reasonable request. The utilized R. codes for the G- and D- study are available in the **Appendix VI**. English and Arabic assessment instruments are available in **Appendix III**.

4. RESULTS

4.1 CHAPTER GUIDE

This chapter is intended to:

- Present generalizability studies' findings.
- Show the exploratory factor analysis results
- Provide additional quantitative findings of the self-assessment and patient assessments.
- Present themes that highlight the impact of the assessment approach.
- Explore themes that reveal how students self-assessed and the factors that influenced their assessments.

4.2 RELIABILITY

4.2.1 Reliability of self-assessment

The G-study for self-assessment (SA) shows that the student (p) main effect is highly significant, accounting for 92.6% of the variance (**Table 4**). The variability of (p) is the desired variability as it denotes true student variance. While the interaction effect of (pc), which is interpreted as case specificity, accounts for 5% of the variance, all other sources have very minor influence (<1%) on the variance individually. The G-coefficient was 0.935, with a standard error of measurement (SEM) of 0.231. A word of caution is warranted, as the rater effect in self-assessment is conflated with the person effect (more about this later in the Discussion section).

Table 4. G-study results for the p (student) × i (item) × c (case) design

Effect (α)	DF	SS	MS	N	$\hat{\sigma}^2(\alpha)$	Percent
p (student)	53	8215	155.01	1	0.78	92.6%
i (item)	30	1798	59.94	31	0.006	0.7%
c (case)	5	139	27.76	6	0.002	0.2%
pi (student x item)	1590	3502	2.2	31	0.008	0.9%
pc (student x case)	265	2278	8.6	6	0.042	5%
ic (item x case)	150	164	1.09	186	3e-5	≈0.0%
pic (unexplained variance)	7950	6480	0.82	186	0.005	0.5%

DF: degrees of freedom, SS: sum of squares, MS: mean square, N=number, $\hat{\sigma}^2(\alpha)$: variance explained.

Given the minor influence of the item specificity component (pi) with 31 items and the significant impact of case specificity (pc), the most effective way to boost reliability is to increase the number of cases rather than items. The D-study results (**Table 5**) show that, with 31 items, increasing the number of cases (n_c) reduces the relative error variance $\hat{\sigma}^2(\delta)$ from 0.28 to 0.05, and raises the G-coefficient from 0.73 to 0.93. With four or more cases, the G-coefficient exceeds 0.90, indicating high reliability, with over 90% of the variance due to true student differences (Bloch & Norman, 2012).

Table 5. D-study results for the p (student) × i (item) × c (case) design, showing the variance for each component and interaction.

n_c	*1*	*2*	*3*	*4*	*5*	**6**	7	8
$\hat{\sigma}^2(p)$	0.78	0.78	0.78	0.78	0.78	**0.78**	0.780	0.780
$\hat{\sigma}^2(i)$	0.006	0.006	0.006	0.006	0.006	**0.006**	0.006	0.006
$\hat{\sigma}^2(c)$	0.011	0.006	0.004	0.003	0.002	**0.002**	0.002	0.001
$\hat{\sigma}^2(pi)$	0.008	0.008	0.008	0.008	0.008	**0.008**	0.008	0.008
$\hat{\sigma}^2(pc)$	0.251	0.126	0.084	0.063	0.050	**0.042**	0.036	0.031
$\hat{\sigma}^2(ic)$	1.7×10^{-4}	9×10^{-5}	6×10^{-5}	4×10^{-5}	3×10^{-5}	$\mathbf{3\times10^{-5}}$	2.4×10^{-5}	2.1×10^{-5}
$\hat{\sigma}^2(pic)$	0.026	0.013	0.009	0.007	0.005	**0.004**	0.004	0.003
$\hat{\sigma}^2(\delta)$	0.285	0.146	0.100	0.077	0.063	**0.054**	0.047	0.042
G-coef	0.732	0.842	0.886	0.910	0.925	**0.936**	0.943	0.949

n_c: number of patient cases. Number of items is fixed (n=31). The bold column shows statistics related to our study specifications.

4.2.2 Reliability of patients' assessment

The G-study findings for patient assessments (**Table 6**) show that the student (p) main effect accounts for the majority of variance (42.6%). The rater (r), which is the patient assessor, main effect and the student-rater interaction (pr), which is interpreted as the assessor severity by the student, also contribute significantly, explaining 33% and 19.5% of the variance, respectively. All other sources of error contribute less than 3% individually. The G-coefficient is 0.66 with a SEM of 0.257.

Table 6. G-study results for the p (student) × i (item) × r (rater) design.

Effect (α)	DF	SS	MS	n	$\hat{\sigma}^2(\alpha)$	Percent
p (student)	53	1940	36.6	1	0.129	42.6%
i (item)	30	2358	78.6	31	0.007	2.3%
r (rater)	5	5144	1028.8	6	0.100	33.0%
pi (student x item)	1590	2165	1.4	31	0.002	0.7%
pr (student x rater)	264	3148	11.9	6	0.059	19.5%
ir (item x rater)	150	1737	11.6	186	0.005	0.3%
Pir (error)	7920	7275	0.9	186	0.005	1.6%

Number of items is fixed at 31.

Given the pi component's minor influence and the pr component's larger impact, increasing the number of raters was identified as the logical strategy to improve reliability. The D-study results (**Table 7**) show that, with 31 items, increasing the number of raters from 1 to 6 dramatically reduces the relative error variance from 0.387 to 0.066, improving the G-coefficient from 0.25 to 0.66. A G-coefficient of 0.72 is achieved with 8 raters, 0.80 with 13 raters, and approximately 0.90 with at least 30 raters. The point of diminishing returns for increasing the number of raters is illustrated in **Figure 7**.

Table 7. D-study results for the p (student) × i (item) × r (rater) design, showing the variance for each component and interaction.

n_c	*1*	*2*	*3*	*4*	*5*	***6***	*8*	*13*	*20*	*30*
$\hat{\sigma}^2(p)$	0.129	0.129	0.129	0.129	0.129	**0.129**	0.129	0.129	0.129	0.129
$\hat{\sigma}^2(i)$	0.007	0.007	0.007	0.007	0.007	**0.007**	0.007	0.007	0.007	0.007
$\hat{\sigma}^2(r)$	0.600	0.300	0.200	0.150	0.120	**0.100**	0.075	0.046	0.030	0.020
$\hat{\sigma}^2(pi)$	0.002	0.002	0.002	0.002	0.002	**0.002**	0.002	0.002	0.002	0.002
$\hat{\sigma}^2(pr)$	0.355	0.177	0.118	0.089	0.071	**0.059**	0.044	0.027	0.018	0.012
$\hat{\sigma}^2(ir)$	0.006	0.003	0.002	0.002	0.001	**0.001**	0.001	0.0005	0.0003	0.0002
$\hat{\sigma}^2(pir)$	0.030	0.015	0.010	0.007	0.006	**0.005**	0.004	0.002	0.001	0.001
$\hat{\sigma}^2(\delta)$	0.387	0.195	0.130	0.098	0.079	**0.066**	0.050	0.032	0.022	0.015
G-coef	0.250	0.399	0.497	0.567	0.620	**0.660**	0.719	0.802	0.857	0.895

n_c: number of cases, G-coef= G-coefficient. The bold column shows statistics related to our study specifications.

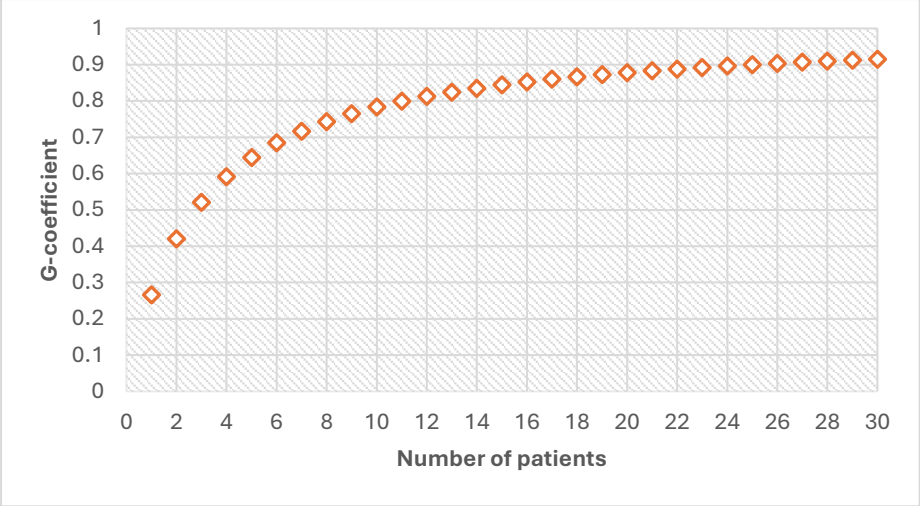


Figure 7. Generalisability coefficients for different numbers of patient assessors in the patient assessment data.

When comparing the reliability of scoring between standardised patients (SPs) and real patients (RPs), the G-coefficient for RPs (0.51) is slightly lower than for SPs (0.55). In both cases, the student (p) is the most significant source of variance (**Table 8**). However, there are notable differences in other sources of variance. For RPs, the student-rater interaction (pr) accounts for 40.6% of the variance, while for SPs, this interaction explains only 26.9%. Additionally, rater (r) variance, interpreted as rater-related error (interpreted as consistency, stringency/leniency, or case difficulty since they are conflated), accounts for 29.4% of the variance with SPs, but only 6.6% with RPs.

Table 8. Comparison of G-studies between standardized patients and real patients (Numbers: person=1, items=30, rater=3,).

Effect (α)		Standardized patients		Real patients	
	n	$\hat{\sigma}^2(\alpha)$	Percent	$\hat{\sigma}^2(\alpha)$	Percent
p (student)	1	0.15	37.8%	0.14	46.1%
i (item)	31	0.007	1.7%	0.008	2.6%
r (rater)	3	0.117	29.4%	0.020	6.6%
pi	31	0.003	0.8%	0.002	0.7%
pr	3	0.107	26.9%	0.123	40.6%
ir	93	0.002	0.5%	0.002	0.7%
pir (error)	93	0.012	3.0%	0.008	2.7%
$\hat{\sigma}^2(\delta)$		0.121		0.133	
SEM		0.345		0.363	
G-coef		0.553		0.512	

The relative error variance $\hat{\sigma}^2(\delta)$ and G-coefficient are calculated at 31 items and 3 raters.

SEM=standard error of measurement.

Figure 8 compares G-coefficients between self-assessment and assessments by real and standardised patients. It shows that while only 2 cases are needed for self-assessment to reach a G-coefficient of 0.80, at least 12 patients (real or standardised) are required to achieve a similar G-coefficient. The reliability of standardised patient assessments is slightly higher than that of real patients, with the difference narrowing to 0.02 by the 12th patient.

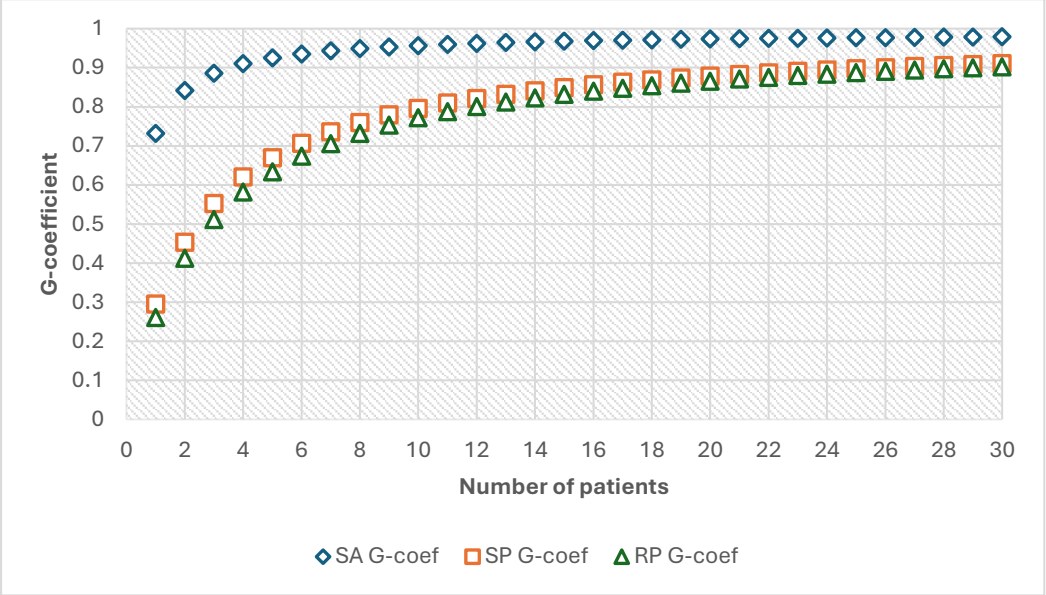


Figure 8. Generalisability coefficients for different numbers of patients (cases for self-assessment and raters for patient assessment) across the three raters (self-assessment [SA], standardized [SP] and real patient [RP] assessments).

4.3 STRUCTURAL VALIDITY

Exploratory factor analysis was conducted on 30 self-assessment items (excluding the global assessment item 31) across six patient encounters. Using Kaiser's criterion (Eigenvalue > 1), four factors were initially extracted (Kaiser, 1970). However, the scree plot (**Figure 9**) indicates a clear inflection point at the second factor, suggesting a single-factor solution (Cattell, 1966). The 4-factor solution (**Table 9**) reveals that the second factor explains less than 5% of the variance, while the third and fourth explain less than 3%. Research indicates that using the 'Eigenvalue > 1' rule can lead to overfactoring (Henson & Roberts, 2006); therefore, some studies suggested using stricter Eigenvalues of >1.5 (Schönrock-Adema et al., 2009); when applied to our study, this leaves 2-factors. Nonetheless, since the additional explained variance accounted for by the second-factor

solution is less than 5%, a one-factor solution was adopted, which converges with the solution suggested by the scree plot (Cattell, 1966; Schönrock-Adema et al., 2009). The one-factor solution explained 52.1% of the total variance and had a coefficient alpha of 0.96.

Table 9. Initial Eigenvalues and total variance explained by extracted factors.

	Initial Eigenvalues			**Extraction SS loadings**		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	16.00	53.33	53.33	15.63	52.11	52.11
2	1.65	5.50	58.83	1.31	4.36	56.47
3	1.26	4.19	63.02	0.89	2.96	59.43
4	1.11	3.68	66.70	0.77	2.58	62.01
5	0.847	2.823	69.525			
6	0.742	2.473	71.998			

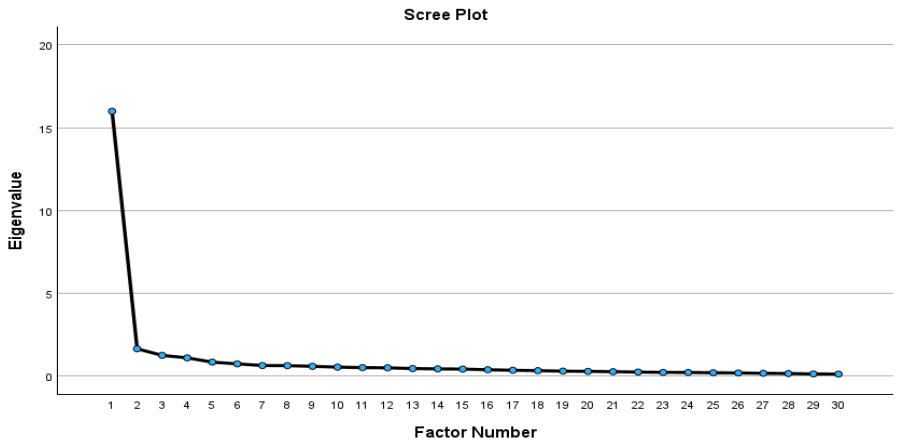


Figure 9. Scree plot for the extracted factors.

Table 10. Descriptive statistics and factor matrix for 324 filled forms (54 participants x 6 examinations).

		Descriptive statistics		Factor solution	
No.	Item	Mean	SD	Factor loadings	h^2
1	Greetings and introducing yourself to the patient appropriately	7.15	1.44	0.76	0.58
2	Identifying the reason for the visit accurately	7.56	1.20	0.76	0.57
3	Clearly outlining what will occur during the session	6.85	1.59	0.74	0.54
4	Employing effective questioning techniques, (using both open and closed questions and justified the use of seemingly unrelated questions where appropriate)	6.79	1.45	0.76	0.58
5	Exploring patients' thoughts, concerns, and emotional state regarding the problem	6.55	1.60	0.73	0.54
6	Requesting consent	7.05	1.59	0.67	0.46
7	Explaining the reasons and steps involved in the clinical examination with respect to the level of information appropriate for the patient (brief, elaborate)	6.81	1.46	0.69	0.48
8	Handling interactions with sensitivity and understanding	7.49	1.28	0.80	0.63
9	Proactively informing the patient about potential discomfort during the examination	6.63	1.81	0.61	0.37
10	Delivering information about diagnosis, causes, mechanism of diseases in a simple manner	7.13	1.31	0.80	0.63
11	Acknowledging deficiency in knowledge or skill (Requests help/time to consult someone when necessary)	6.79	1.61	0.56	0.31
12	Discussing treatment options, including costs, benefits, and risks, offering choice as applicable	6.55	1.79	0.61	0.37
13	Providing preventive and treatment instructions	6.97	1.78	0.57	0.32
14	Motivating the patient to adhere to the agreed-upon treatment plan (giving motivating reasons, encouraging benefits, awareness of culture, empowers the patient)	6.69	1.68	0.64	0.42
15	Summarising key points in the session	6.90	1.54	0.74	0.54
16	Providing a clear plan for the next session	6.79	1.66	0.70	0.49
17	Appropriately addressing any remaining questions, concerns, or requests (relaying any remaining questions to the next session if time does not allow)	7.22	1.29	0.73	0.54
18	Providing safety nets (what to do in case things went in an unpleasant direction)	6.25	1.77	0.64	0.41
19	Treating the patient with respect	7.98	1.03	0.75	0.57
20	Demonstrating active listening	7.92	1.07	0.72	0.52
21	Demonstrating genuine care and concern for you as a person and responds with empathy	7.67	1.14	0.79	0.62
22	Encouraging the patient to ask questions about anything.	7.58	1.26	0.72	0.52
23	Displaying humility in interacting with the patient	7.78	1.06	0.74	0.55
24	Using appropriate verbal and nonverbal communication effectively	7.04	1.40	0.81	0.66
25	Inspiring trust/reassurance	7.15	1.38	0.82	0.67
26	Organizing the interview well (time, sequence of events, systematic, signposting)	6.68	1.63	0.77	0.60
27	Personalizing the session to the patient (not robotic).	7.02	1.25	0.75	0.57
28	Have a professional appearance	7.58	1.16	0.62	0.39
29	Ensuring shared understanding throughout the session	7.15	1.28	0.76	0.58
30	Encouraging patient's involvement and shared decision making according with patients' preference	6.82	1.44	0.71	0.51

Note: h^2= communality

To check the reliability of the EFA output, I split the data randomly in half and conducted EFA for each half. Almost the same Eigenvalues and scree plots were produced. Using the same factor criteria above, the one-factor solution was the most appropriate for either half, the same as the factor solution for the entire dataset.

A total of 28 items (out of 30) had loadings larger than 0.60 (**Table 10**). Only items no. 11 and 13 had loadings less than 0.60 (0.56 and 0.57 respectively). Therefore, this one factor can be considered reliable regardless of the sample size (Guadagnoli & Velicer, 1988)Since this assessment instrument was designed specifically to assess communication skills (see section 2.3), the interpretability of the single extracted factor is theoretically supported.

Rotated factor loadings for a two-factor solution are shown in **Appendix IX**.

4.4 FURTHER QUANTITATIVE ANALYSIS

Since exploratory factor analysis indicated that all items correspond to a single factor, aggregating scores across items is justified. Therefore, both SA and PA scores were aggregated (summed) for the 30 items. The normality of the aggregated scores was checked using the Kolmogorov-Smirnov test, which was not significant ($P > 0.001$), while the Shapiro-Wilk test was significant ($P < 0.001$). However, skewness and kurtosis values were within acceptable ranges for parametric tests, with thresholds of ±2.0 for skewness and ±7.0 for kurtosis, following Hair Jr et al. (2010) guidelines. Detailed test results are provided in **Appendix X**. Therefore, the measures were treated as normal, applying parametric statistics.

Descriptive statistics for SA and PA per patient ID are shown in **Table 11**, along with ANOVA results. PA scores for real patients (RPs) were significantly higher than those for standardised patients (SPs), with a very large effect size ($P < 0.001$, $d = 1.51$), according to an independent sample t-test. No significant differences were found in SA scores per patient ID ($P = 0.52$, $\eta^2 = 0.01$).

A mixed repeated measures ANOVA revealed consistent improvement over time for both SA and PA (**Figure 10**), with significant differences and large effect sizes (**Table 12**). The order in which students examined patients varied, so changes in PA mean scores cannot be attributed to patient ID. Within-subject effects, including sex, family income, and

social activity per week, were insignificant for both SA and PA. However, the social type (introvert/extrovert) had a minimally significant effect on SA ($P = 0.08$) and a significant effect on PA ($P = 0.046$). No between-subject effects were significant, but large effect sizes were observed for sex and social type in SA ($P < 0.10$, Partial $\eta^2 > 0.14$).

Table 11. Descriptive statistics for the aggregated self-assessment and patient assessment as per patient ID with One-way ANOVA test.

		Self-assessment		Patient assessment	
Patient ID	N	Mean	SD	Mean	SD
Lama (SP)	54	212.1	28.9	181.5	26.5
Mohy (SP)	54	213.4	30.9	211.7	11.5
Bashar (SP)	54	208.3	34.3	213.8	22.0
Yara (RP)	54	208.0	31.9	247.4	16.1
Layla (RP)	54	216.8	30.8	230.4	18.0
Sara (RP)	54	216.6	29.3	238.6	28.7
One-way ANOVA					
P-value		0.529		<0.001	
η^2		0.013		0.511	

SP=standardised patient; RP: real patient. η^2=eta square

Table 12. Mixed repeated measures ANOVA statistics for aggregated self-assessment and patient assessment as per patient order.

		Self-assessment		Patient assessment	
	N.	Mean	SD	Mean	SD
1st patient	54	198.9	31.2	203.1	30.1
2nd patient	54	211.9	28.6	225.1	24.2
3rd patient	54	214.9	33.0	222.9	26.2
4th patient	54	216.6	33.9	221.9	31.8
5th patient	54	216.8	28.7	222.8	32.2
6th patient	54	216.1	28	227.5	30.8
Within-subject effects		**P-value**	**Partial η^2**	**P-value**	**Partial η^2**
Time*		<0.001	0.184	<0.001	0.131
Time × sex		0.551	0.014	0.123	0.045
Time × family income		0.084	0.158	0.079	0.114
Time × social type		0.081	0.075	0.046	0.06
Time × social activity per week		0.359	0.06	0.415	0.056
Between-subject effects		**P-value**	**Partial η^2**	**P-value**	**Partial η^2**
Sex		0.056	0.198	0.148	0.119
Family income		0.541	0.116	0.915	0.029
Social type		0.092	0.158	0.232	0.083
Social activity per week		0.333	0.121	0.755	0.033

*Sphericity assumption violated; Greenhouse-Geisser correction applied.

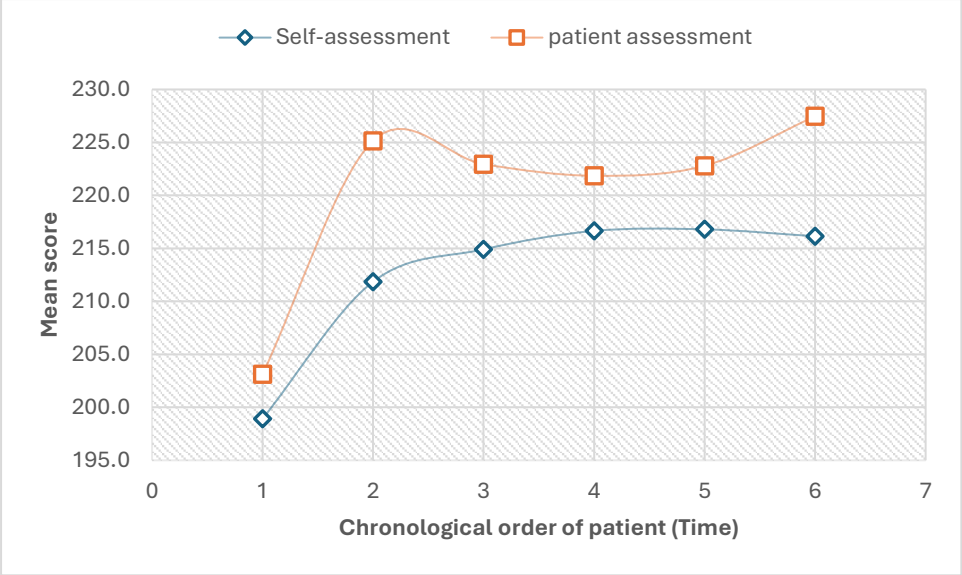


Figure 10. Aggregated patient assessment and self-assessment as per the chronological order of patient consultation (Time)

Table 13 shows Pearson correlations between SA, PA, university grade, the global SA item, the global PA item, and the initial global self-assessment (a 9-point item on communication skills assessed before the assessment sessions began). All correlations, except between PA and initial global SA, were significant, though most were small ($0.2 < r < 0.29$) according to Cohen (1988, pp. 79-81). Notably, strong correlations were observed between SA and the global SA item and between PA and the global PA item, indicating that global items correlate well with aggregated analytic items. The correlation between PA and university grade was very small ($r=0.155$).

A mixed repeated measures ANOVA for comparing the three PA tiers is presented in **Appendix XI**. The large number of students examining them could have affected real patient assessment due repeated exposure to dental and disease terminology. To ensure consistency, an independent t-test was conducted comparing RPs assessments of the first and last 22 students, revealing no significant difference ($P=0.4$, $d=0.13$).

Table 13. Pearson correlation and confidence intervals between Self-assessment, patient assessment, university grade, initial global SA (assessed before approaching patients) and the global SA item (assessed during sessions for each patient)

		N.	r (95% CI)[a]	*P*-value
SA	Uni grade[b]	306	0.260 (0.15-0.36)	<0.001
	Initial global SA	324	0.247 (0.14-0.34)	<0.001
	Global SA item	324	0.794 (0.75-0.83)	<0.001
PA	SA	324	0.234 (0.12- 0.33)	<0.001
	Uni grade	306	0.155 (0.04-0.26)	0.007
	Initial global SA	324	0.069 (-0.04 -0.17)	0.215
	Global SA item	324	0.314 (0.21-0.41)	<0.001
	Global PA item	324	0.856 (0.82-0.88)	<0.001

a Estimation is based on Fisher's r-to-z transformation with bias adjustment.
b three students did not report their university grade
Note: Uni grade %=the average cumulative university grade. SA=self-assessment,
PA=patient assessment. Three students did not specify their uni grade which resulted in the loss of 18 data points.

4.5 IMPACT OF THE ASSESSMENT APPROACH

This section presents thematic analysis findings regarding the impact of the assessment approach. **Table 14** shows the characteristics and mean SA and PA for the six encounters of Student interviewees. Interviewees will be mentioned in the quotes, and their PA and SA scores are important to understand their experience.

Three main impact-related sub-themes were generated: Self-control, Now I See It!, and Self-assessment versus Patient assessment.

Table 14. Interview student participants' characteristics and mean self- and patient assessment for the six encounters.

Pseudonym	Sex	Uni grade[a]	PA tier	SA tier	Initial Global SA[b]	Patient assessment		Self-assessment	
					1 to 9	Mean	SD	Mean	SD
Hassan	Male	78	Low	High	7	212.8	34.2	244.3	29.3
Maher	Male	80	Low	High	7	216.3	23.7	236.0	17.0
Ahmad	Male	83	Low	High	7	210.2	37.2	258.7	6.7
Bahaa	Male		Low	Low	4	204.8	46.0	149.3	39.0
Rama	Female	76	Low	Low	4	204.8	21.5	186.0	10.6
Sham	Female	79	Low	Low	6	210.0	45.6	191.0	7.3
Abd	Male	85	High	high	6	238.5	23.9	230.0	11.6
Khattab	Male	85	High	high	7	236.2	36.2	227.5	12.1
Mariam	Female	89	High	high	9	236.2	30.8	244.7	3.1
Rania	Female	89	High	Low	7	228.7	27.9	178.7	8.2
Fatima	Female	79	High	Mid	3	232.3	10.3	198.3	15.2
Shadi	Male	86	High	Mid	6	247.8	22.4	214.0	4.5

a University grade is calculated out of 100. b.Initial global SA =general self-assessment of communication skills prior to patient examination. PA=patient assessment.

4.5.1 Self-control

This sub-theme, rooted in Zimmerman's SRL theory, pertains to self-control during performance, which helps learners focus and optimise their tasks (Zimmerman, 2000). Interviewees expressed that engaging in self-assessment (SA) facilitated self-control, manifesting through self-instruction and attention focusing. Self-instruction is covertly or overtly describing how one performs a certain task (Zimmerman, 2000). Interviewees

expressed that being involved in SA helped them develop a kind of "*algorithm*" of how to start, what to do and say during, and how to conclude, i.e. it gave them structure and instructions. Many interviewees echoed this idea, and it is illustrated well in Rania's following account when she was asked to explain her improved SA in the last patient:

Rania: when I got to the last patient, I got the hang of it, you understand? I got what the assessment form requires….. I organized what I need to do before each patient: I will do this, this and this; these are the things that I usually forget to do based on my previous self-assessment.

Engaging in self-assessment can also work as a persistent reminder of what needs to be done, even in difficult cases such as Rama's:

Interviewer: Did you use to deal with patients in a manner where all items are performed and done well?

Rama: This is something I noticed everyone was doing but I did not. Preventive instructions, a close example, I did not give preventive instructions though it was clear in the form…in normal daily life, I rarely give preventive instructions..

Interviewer: …but in the last patient, you provided preventive instructions (patient's assessment was 9/9 on this item)

Rama: Correct, she is the only patient, at the end, after each patient, when I get to this item (during SA), I would say 'again, I did not give preventive instructions', okay, the next patient, then I forget…until the last time, I told myself I will give preventive instructions, so I did the last time.

Rama consistently attempted self-instructions after each SA but often forgot until the final patient consultation, which she performed successfully on a different day. Notably, she had reviewed previous patients' assessments before attempting the last consultation which was on a different day; she acknowledged that this also could have influenced her behavior change.

Self-control was also evident through focused attention. Interviewees reported increased concentration on areas where they previously scored poorly. For instance, Shadi noted improved motivation, concentration, and care:

Shadi: What I felt the most is how better I need to be with the next patient, let's say it's a kind of self-challenge. With each patient, with this patient I focused on this, this and this and I was good at this and this, but here, I did not do well, here I did not self-assess well. So in the next patient, I need to be more careful and more aware.

4.5.2 Now I see it!

This sub-theme is characterized by participants gaining new insights into their communication performance and identifying strengths and weaknesses through the assessment approach, particularly self-assessment. Participants reported increased self-awareness regarding their strengths and areas for improvement. For example, Sham realized she was not introducing herself effectively at the beginning of sessions:

Sham: The form structured the session, split it into parts, after this splitting, I started to see where my weaknesses are, I felt that my weakness point is introducing myself and talking more with the patient.

This inference was concurrent with patients' assessment.

Sometimes, when blind spots remained blind even after SA, cross-analysis of SA with PA was essential to reveal them. For instance, Hassan articulated he were not aware that oral examination could cause discomfort to the patient until they were informed of PA. Hassan had a high SA but low PA; his high self-assessment likely reflects high self-confidence, which is evident in the interview, which may have blinded him to areas needing improvement. In contrast, Bahaa, with both low SA and low PA, initially rated himself 2/9 on forewarning the patient but improved to 8/9 with the last patient (patient assessment was 9). Bahaa explained his initial low rating and demonstrated his awareness of patient discomfort:

Bahaa: I became aware of it later with the last patient I examined (forewarning the patient)..maybe the dental mirror, moved a little closer to his pharynx, he had a gag reflex which bothered him. Maybe that's why I put 2.

Interviewer: What made you forewarn the patient (last one)? What made you change?

Bahaa: The form had a role in alerting me, and the incident that happened to me….

4.5.3 Self-assessment versus patient assessment

This theme explores the perceived validity of SA compared to PA. Most participants viewed PA as more valid, though some believed their SA was more accurate. When PA was lower than SA, some attributed it to the patient's strictness or difficulty, a perception that spanned across all PA tiers. What is that they were right at some cases.

Participants suggested that both SA and PA should be used together, followed by a dialogue to prevent being "wronged." Sham's account illustrates this point:

Sham:…it is important that I self-assess

Interviewer: Why?

Sham:..a patient wrote that I did not explain medical terminology though I did… I feel like if it is only me who assesses, I could wrong the patient, and if it is just the patient who assesses, they could wrong me. Do both assessments; then after we are done, we put it to the discussion.

On the other hand, where SA was lower than PA, Fatima expressed that SA is more "accurate" than PA because she is more honest with herself. Shadi also expressed that patients are not always aware of their rights, especially in the Syrian culture, so they would assess you well on something like 'asking for consent' because these practices are not common and are not expected of the dentist in the first place.

In contrast, Mariam, in the following account, expressed that their SA does not matter because of its unreliability and that what really matters is PA:

Mariam:…my self-assessment will not change, self-assessing myself won't change the outcome…. a person will definitely see himself in perfect image, will always feel that he did everything well, that he is always in the front, and better than others; patient assessment is what matters at the end.

Mariam's initial global SA (before approaching patients) was 9/9, and she consistently self-assessed herself higher than most patients did assess her (**Table 14**); this might explain where she is coming from when making her aforementioned point. However, it is important to note that Mariam did actually talk about a couple of points where her self-assessment improved her performance, such as reminding her of tasks that she usually forgets or did not consider doing before.

4.6 SELF-ASSESSMENT UNDER THE MICROSCOPE: EXPLORING JUSTIFICATIONS

This section delves into interviewees justifications of self-assessment scoring and reasons behind the alignment or misalignment with PA, focusing on three emergent themes: subjective use of the assessment form, patient-dentist dynamics, and self-reflection.

4.6.1 Subjective use of the assessment form

Students interacted subjectively with the self-assessment form, interpreting items and using the 9-point scale in varied ways. While some rated themselves 6 or 7 for good performance, others favoured 8 or 9. A narrow scoring range (5-7) was adopted by some, whereas others consistently chose the upper range (7-9). This contrasts with the broader range of scores given by patient assessors. Rama's account, echoed by many participants, illustrates this point:

Rama: I used to assess 5 when I was clearly bad, 7 clearly good, and 6 neutral.

Generally, the use of scores 1-2-3 was very rare, even when a task was not performed entirely; they would self-assess 4 or 5 instead. Some participants did not have a reason for why they did that. Nevertheless, two participants (Rama and Abd) said that self-assessing in the lower range would upset them. Here is Abd's account of this:

Abd: There is a patient to who I forgot to give preventive instructions at all
Interviewer: You self-assessed 5.. Why didn't you put 3, 2, or 1 instead?

Abd: I don't know. Should I put 1 instead?

Interviewer: Logically, one puts 1, 2, 3…

Abd: I find 5 average, same as nothing

Interviewer: Could it be….what would happen if you self-assessed 3 or 2?

Abd: Honestly, I would be a bit upset (laughing)

Some students expressed that they like scoring in the middle, neither high nor low, such as in the case of Rania; here is her account explaining why she did not use a wider range of scores:

Rania: Absolutely. I should have used a wider range, just like I told you…As a personality type, I like being in the middle, maybe a bit more, maybe a bit less. There is no WOW, excellent or something very low unless I really felt that.

It was also observed that interviewees were assigning specific meanings to grades as if they had a grading rubric in their minds. For instance, Shadi explains clearly the difference between 7 and 8 for him when self-assessing the "*asking for consent*" task:

Shadi: When I put 8, it means that I told the patient exactly what I would do with the dental examination tools and whether he would allow me to use them or not. I told him that I would use the dental mirror inside the mouth, retract the cheeks, and retract the lip. When I say these things, I self-assess 8 because then I would be confident that things are okay, and I asked for his consent even to cause discomfort.

Conversely, students like Hassan, Maher, and Ahmad, who used simple yes/no interpretations of items, tended to have a limited range of scores and lower patient assessments (PA). Those who provided more nuanced interpretations of what constituted good execution of an item used a wider range of scores and had higher PA.

Furthermore, participants sometimes connected assessments of different items together, i.e., the score on a certain item was affected by general performance. For example, Fatima self-assessed 3 and another time 5 in forewarning the patient task, though, in both instances, she knew that she had not done the task. When asked to explain why, she answered:

Fatima: Yeah, this is probably psychological (laughing)…with the first patient, I wasn't focused, so I put low scores across all (items)..whereas here, the overall performance improved, so yeah, why not (put high scores)…. I did not isolate the item from others; no, all are similar to each other, you know? obviously, I'm not going to put 8 in one and 1 in another; I tried to make them close relative to the general performance.

An important sub-category that emerged across all participants is *external referencing*, in which students judge their ability according to external criteria. Many participants referenced their ability to diagnose or plan treatment explicitly when asked to justify their self-assessment, though the items were about communication skills.

4.6.2 Patient-dentist dynamics

Student participants identified several factors related to patient-student dynamics that influenced their self-assessment. Many reported deliberately rating themselves lower to account for the possibility of receiving a lower rating from the patient. For instance, Sham consistently rated herself 7 in the "listening attentively" task, while all patients rated her 9, explaining:

Sham: I don't know, I would not self-assess 9. There is the possibility that the patient did not feel that I heard him attentively, so I won't put 9 because I'm not sure of the patient.

Participants also tended to rate themselves higher when they perceived the patient as satisfied, using facial expressions, tone, and behavior as cues. However, the level of patient interaction had the most significant impact on SA. Many students rated themselves lower when patients were unresponsive and higher when patients were engaged. Rania, for example, self-assessed lower with a patient who seemed disinterested, saying:

Rania: What happened is, when I'm talking and explaining in length, I felt she was not interested.., I felt something telling me 'That's enough, stop!' or what I need to do to make her more interactive (thinking out loud); that's why I self-assessed lower.

Chemistry with the patient was echoed by many participants as a reason why they self-assessed higher at times. Fatima rated herself 8/9 for the first time in employing effective questioning techniques, attributing it to a positive feeling she had about the patient even before the consultation began.

Fatima: First thing is, the same patient, before I talked to her, I saw her from afar. I felt she has a sweet personality without any specific reason, I felt that I will be comfortable with her, she won't speak ill of me…she won't judge me.

Gender and age concordance further influenced students' communicative performance and SA, with male students sometimes distancing themselves from young female patients and female students feeling more comfortable with female patients.

Interpersonal sensitivity also impacted SA scores. Students more attuned to patients' discomfort tended to report it more accurately and rated themselves lower on related items, unlike those less aware. Maher's account illustrates this:

Maher: *This is the patient who spun his head a little, and I wasn't paying attention...he was bothered by the mirror… He told me that afterward…. I felt he was a bit too whiny. I examined all patients the same way. If I bothered him, two or three out of six would have reported it. I don't know, I didn't feel I bothered him.*

It was important to note that most participants did actually have similar complaints and this was revealed to Maher in a friendly manner, who appreciated the feedback.

4.6.3 Self-reflection

Self-reflection has been defined in various ways in the literature; from a social perspective, it is described as "*a temporary phenomenological experience in which the self becomes an object to oneself*" (Gillespie, 2018, p. 245). In this study, several subthemes related to self-reflection were identified, including self-observation, self-criticality, self-reaction, and self-concept, all of which influenced self-assessment.

Self-observation was demonstrated as all students correctly identified some areas where they did well and areas that needed improvement.. A case in point is the following account with Rama, who was answering why she self-assessed 5 on the introducing oneself (item 1):

Rama: According to what I remember, when I introduced myself to this patient, I said 'Hi, I'm Rama'; then he said, Dr. Rama, I felt it was wrong to introduce myself as Rama, as he is coming to me for treatment; I'm Dr. Rama.

According to the session recordings, Rama changed her self-introductions and introduced herself as Dr. Rama with the following patient.

Strengths were generally easier to assess than weaknesses, and some participants struggled to recognize their mistakes even after reviewing patients' assessments.

Self-criticality was observed in all groups except the low PA - high SA tier. Participants held themselves to higher standards, acknowledged the possibility of error, and recognized the need for further learning, or directly criticizing their performance. Shadi, for example, self-assessed lower on asking for consent despite being rated 9/9 by patients because he believed he could have done more:

Shadi: ..deep inside, I knew I could ask them more. For example, at the end, during examination, we do not only use examination instruments, we are wearing gloves, so our

hands our sterile, so we can put our fingers inside to do teeth percussion or examination or to feel them. I should have, in my opinion, asked for consent for this thing….I did not do it, so I self-assessed lower.

Rania also mentioned that she thinks she can still do better, and that's why she generally assessed herself lower than patients' assessed her:

Rania: …I say to myself, I can do better, I can do better, I can do better. That's why I do not feel I reached the end of the road…

Self-reaction, which describes how participants felt about their performance, also affected SA. This influence was either direct or indirect. For instance, Khattab expressed being "very satisfied" with his last patient consultation, which led to an increase in his SA. Similarly, Rama's dissatisfaction with her first consultation led her to under-assess herself with subsequent patients.

Self-concept was an important sub-theme that affected SA, with some students identifying themselves as good communicators and others on the opposite end. The initial global self-assessment measure reflected these perceptions (**Table 14**) which is also concordant with the thematic analysis. An example where a participant showed a high self-concept of is Hassan:

Hassan: ..the person needs to have high confidence in himself, he must always consider himself the best so that he achieves success and improves himself, you see?

Maybe this assumed self-concept explains why he had high SA, though PA was on the lower end. In contrast, Bahaa said when asked about why he self-assessed on the low end of the scale in the "*Initiating the session and establishing rapport*" whole subsection:

Bahaa: To start, I'm not very good at communication skills, which is opposite to what our profession requires. That is life. I prefer being an introverted and isolated person, I do not have a lot of friendships, I'm not a very social person.

5. DISCUSSION AND CONCLUSIONS

5.1 CHAPTER GUIDE

This chapter is intended to:

- Draw inferences from findings in relation to the research questions
- Synthesize findings from both the quantitative and qualitative strands to address the impact of the assessment approach
- Discuss and interpret findings within relevant academic literature
- Highlight the study's original contribution to relevant literature
- Analyse the strengths and limitations of the study
- Identify directions for future studies
- Discuss implications for educational practice

5.2 VALIDITY OF THE ASSESSMENT APPROACH

5.2.1 Content and ecological validity

This study aimed to develop and validate a self-assessment approach for communication skills in clinical settings. The assessment instrument was based on the Calgary Cambridge Guide (CCG), with items systematically sampled and supported for relevance in dental education, thereby supporting the instrument's content validity (Linn, 1993, p. 65). A robust cross-cultural validation process, following Sousa and Rojjanasrirat (2011) guidelines supported the instrument's relevance for assessing communication skills in dental consultations in Syria.

Assessments were conducted in authentic clinical settings, using real examination tools and diagnostic images. Students examined both standardised and real patients, which were often indistinguishable, according to interviewees. The use of authentic elements in a real-world context lends high ecological validity to the assessment (Govaerts & van der Vleuten, 2013; Phillips, 2014).

5.2.2 Reliability of the assessment approach

Reliability in this study was assessed using generalizability theory, which expands on classical test theory by measuring errors from multiple sources (see Section 3.5.4). The study found that student-related variance accounted for 92.6% of the total variance, with a

G-coefficient of 93.5. This indicates that 93.5% of the observed score variance is attributable to true differences among students and/or raters, as these are conflated.

An extensive literature review revealed no previous studies calculating the G-coefficient for self-assessment of communication skills, which is understandable given that G-studies are usually conducted within high-stake objective structured clinical examinations (OSCE) with no self-assessment component (Bloch & Norman, 2012; Monteiro et al., 2019). However, a survey study by Peterson et al. (2012) found that self-assessment of medical competencies could achieve a G-coefficient over 0.9 with fewer raters than required for faculty assessments, aligning with our findings.

Several factors may explain the high G-coefficient for self-assessment. First, the student effect is conflated with the rater effect, as the rater is the student (self-assessment), which may inflate the variance attributed to students. Consequently, the exact student-attributed variance cannot be isolated, requiring a cautious interpretation of the G-coefficient. Additionally, qualitative interviews suggests that students may conceptualize communication skills differently, often making a global self-assessment and then rating themselves high or low across all items, irrespective of their specificity which could also account for the high G-coefficient (Zanetti et al., 2010).

The G-coefficient was similar for both standardised patients (SPs) and real patients (RPs). However, the smaller rater component variance in RPs suggests they were more consistent than SPs. Conversely, the pr component, which reflects "*whether the rank ordering of the examinees changes depending on the raters*" (Zanetti et al., 2010, p. 277), was higher in RPs (40.6%) compared to SPs (26.9%). This indicates that SPs ranked students more consistently. These findings align with previous research comparing SPs and lay raters of videotaped student encounters (Zanetti et al., 2010). The results also show that RPs were more lenient and tended to give uniformly high ratings, possibly making global assessments that overlooked item specificity. In contrast, SPs were more likely to assess each item individually, adjusting scores significantly for specific tasks even in high-performing students. Despite these differences, it's reassuring that student-related variance was the largest component in both SP and RP assessments.

A meta-analysis of generalizability studies in technical skills assessment reported a pooled participant effect of 44.2% (95% CI, 36.8%–51.5%) and a pooled G-coefficient of 0.65 (95% CI, 0.59–0.70) (Andersen et al., 2021). Although the current study focused on

communication skills, which are inherently more subjective and challenging to assess than technical skills (Claramita et al., 2013; Kruse et al., 2020; Röttele et al., 2021; Verheijden et al., 2023), the G-coefficient from patient assessments aligned with the meta-analysis.

The current study outperformed most generalizability studies on communication skills, with higher student variance in patient assessments **(Table 15)**. Only three studies reported higher G-coefficients, two of which using video or audio recordings rather than direct observation (Gobat et al., 2015; Karabilgin et al., 2012; Zanetti et al., 2010); Live performance assessments, as Andersen et al. (2021) noted, require more observations for reliability. Moreover, using trained external examiners instead of standardised patients might have influenced the reliability of these assessments.

Table 15. Summary of previous G-studies in the assessment of communication skills.

Study	Type of observe.	Type of raters	Number of				G-study		
			Person	Rater	Item	station	Design	P (var%)	G-coef
Hodges et al. (1996).	Live observ.	SP-examiner (paired)	96	4	7	4	-	-	0.17
	Live observ.	SP-examiner (paired)	95	6		6	-	-	0.28
Röttele et al. (2021)	Survey (retrospective)	Self- and real patients	32	284	21	-	P x r x subscale	1%	0.06
Zanetti et al. (2010)	Video rec.	Examiner	20	3	21	4	P x i x r	20%	0.75
	Video rec.	SPs	20	3	21	4	P x i x r	5%	0.53
	Video rec.	Lay raters	20	3	21	4	P x i x r	15%	0.68
Silverman et al. (2011)	Live observ.	Examiner	124	4	15	4	P x r:s	20%	0.5
Karabilgin et al. (2012)	Live observ.	Examiner	1026	7	13	1	P x r x i x s[a]	23.4%	0.836
Gobat et al. (2015)	Audio rec.	Examiner	26	3	5	3	-	-	0.844
Shibli-Rahhal and Kreiter (2021)	Live observ.	SP-companion	140	2	25	1	P x r x i[b]	14%	0.146
	Live observ.	SP-companion	140	5	-	5	P x s		0.50
Daniels and Harley (2017)	Live observ.	Examiner	61	1	4	4	P x i:o[c]	-	0.507
	Live observ.	Examiner	25	1	4	5	P x i:o[c]	-	0.633

P=person, r=rater, i=item, s=station. P (var%)=variance (%) attributed to the person. Live observ.=live observation. SP=standardised patient.
a. Not all raters in this study assessed all students: each rater assessed a selected number of students per year, making the fully crossed design unsuitable. Raters should be nested within students
b. 11 different SPs were forming 28 unique pairings; thus, this is not a fully crossed design but should be rather a P: R x i
c. The g-coefficient was not stated in the study; however, the variance of each component was mentioned, which enabled the calculation of the G-coefficient using equations mentioned in one study (Huebner & Lucht, 2019).

5.2.3 Structural validity of the assessment tool

Using evidence-based EFA methods, this study found that the 30 items of the assessment tool coalesced into a single communication skills factor among clinical dental students in Damascus. Simmenroth-Nayda et al. (2014) reported a 5-factor structure for a German version of the Calgary Cambridge guide, based on a very small sample of 30 participants without reporting the KMO index, eigenvalues, or scree plot. One factor had only one item, and two others had 3 or 2 items, which is problematic as a minimum of three items per factor is generally recommended (Schönrock-Adema et al., 2009). It was also apparent that a single factor accounted for the majority of the variance in their study.

Another study that used an 8-item instrument to assess communication skills showed a good fit for a single factor using confirmatory factor analysis (CFA), though CFA forces items into one factor (Gärtner et al., 2022). Côté et al. (2001) explored the structural

validity of a 20-item instrument for assessing communication skills in Canada. EFA showed a 2-factor solution using eigenvalue>1 as criteria, though the majority of variance was accounted for by one factor.

In a survey study that reported self-assessment of communication skills for medical residents in the US, a two-factor solution was reported explaining 51.4% of the total variance; the factors were called interpersonal relations and conveying medical information; however, ten items out of 11 loaded significantly on both factors. Unfortunately, neither the scree plot nor the eigenvalues were reported to examine how appropriate their factor solution was (Symons et al., 2009).

Despite inconsistent EFA reporting in previous studies, evidence suggests that instruments assessing communication skills often have a unidimensional factor structure, similar to our findings (Côté et al., 2001; Gärtner et al., 2022; Symons et al., 2009). Regardless, the value of this study's single-factor instrument should be assessed based on its meaningfulness, interpretability and consistency across different samples, methods, and studies (Gorsuch, 1983; Tucker & MacCallum, 1997).

Having discussed the evidence for test interpretability, the next section will address an important facet of validity: the educational impact of assessment.

5.2.4 Validity as impact: Integrated conclusions from quantitative and qualitative strands

Evaluating validity in relation to whether the assessment does what it is employed to do requires an appraisal of the potential and actual social consequences (Linn, 1993). This study's assessment approach was designed to facilitate performance improvement and foster meaningful dialogue with external assessments. The quantitative findings revealed significant, consistent improvements in students' communication skills, as assessed by patients. Qualitative data corroborated these results, showing that students used self-assessment to identify areas for improvement, set them as goals and employed self-instruction and focused attention to enhance their communication behavior. The assessment items served as actionable performance goals, motivating students to improve and set higher standards for themselves. It is also apparent in the interviews that students were more motivated to improve so that they could assess themselves better.

Contrary to the common criticism of self-assessment accuracy (Davis et al., 2006; Eva & Regehr, 2008; Yates et al., 2022), cross-analysis of interviews and patient assessment data indicated that most students accurately identified some areas for improvement through self-assessment. This holds true even considering that this study's statistical correlation between SA and PA was low. Students' often rated themselves lower on tasks they recognized as poorly executed. However, what a "low" score is differs between patients and students, and within students themselves as I will explore later.

The combined quantitative and qualitative findings demonstrate that self-assessment drives learning through multiple mechanisms. Self-regulated learning theory offers a useful framework for understanding how SA promotes performance improvements (Zimmerman, 2000). SRL model comprises three cyclical phases are interconnected, with changes in one affecting the others. In this study, systematic self-assessment, influenced students' goal-setting and motivation in the forethought phase and enhanced self-instruction, attention focusing, and self-observation in the performance phase. This led to self-driven improvement facilitated by analytic self-assessment.

Self-determination theory further supports these findings, suggesting that actions performed with intrinsic motivation, driven by a sense of volition and agency, result in better and more sustainable educational outcomes (Williams et al., 1999). The incorporation of patient assessments alongside self-assessment was also crucial; students often viewed patient feedback as a more valid measure of their performance. The cross-analysis of SA and PA revealed important insights that students might have otherwise overlooked.

This study revealed that self-assessment adds a new dimension to external assessment interpretation, resulting in four distinct student groups (**Figure 11**). The assumption that SA is always inflated may lead educators to overlook key insights (Eva & Regehr, 2008). Interviews showed that students in areas B and C, who often underscored themselves, displayed more improvement-oriented actions and deeper reflection than those in areas A and D. This challenges the emphasis on high correlation between SA and external assessments, suggesting that fostering humble, self-critical SA may be more beneficial for learning (Brown et al., 2015; Davis et al., 2006). High SA, even when confirmed, could lead to complacency and superficial reflection, emphasizing the importance of SA as an

indicator of learning behavior and the potential impact of this study's assessment approach (Bowen et al., 2019; Eraut, 2000).

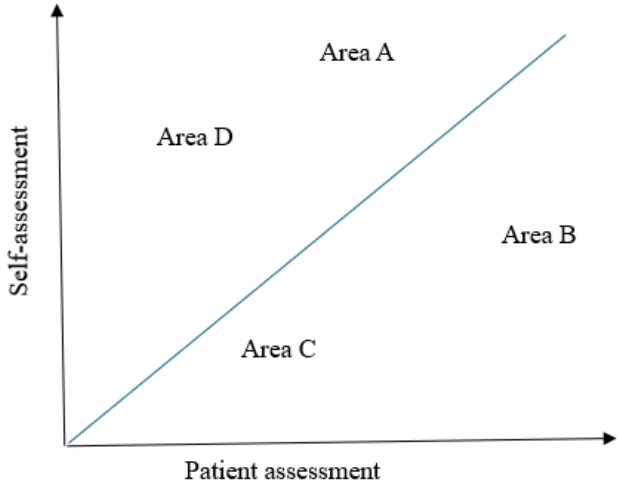


Figure 11. Patient assessment is represented as a line that is intersected with a self-assessment line. The blue line is when self-assessment is equal to patient assessment. The areas above the blue line (A and D)represent those with higher SA and lower PA. Areas below the blue line (B and C) represent students with lower SA and higher PA.

5.3 WHAT DO THESE SELF-GIVEN SCORES MEAN?

Research findings revealed that self-assessment of communication skills is multifactorial, with scores reflecting subjective interpretations. Students varied in their use of the assessment form, with some applying a narrow range of scores while others a broader one. They also assigned specific meanings to scores, such as 5 for "bad" and 7 for "good," diverging from the intended scale use. This variation underscores the importance of understanding how participants interpret scores before statistically assessing self-assessment accuracy, as noted by Ward et al. (2002, p. 70) who showed how "*inconsistent use of the scale among students attenuates the correlation of expert and self-ratings."*

The study also highlighted the influence of the halo effect, where a positive or negative impression in one area impacts unrelated ratings (Nisbett & Wilson, 1977). Additionally, factors unrelated to communication skills assessment, like diagnosis and treatment planning, influenced some self-ratings, suggesting future research should consider including these domains in assessments so that they can assess each separately. It is

important to note that the halo effect, range restrictions and leniency do not just affect self-assessment; they are also common rater errors (Govaerts et al., 2007).

The dynamic between students and patients was another key factor, with perceived patient satisfaction and interaction levels significantly impacting self-assessment. However, reliance on patient satisfaction alone may lead to inaccurate self-assessments, as seen when students rated themselves lower with difficult standardised patients despite high-performance ratings. This indicates that students might rely too heavily on perceived patient satisfaction without critically analysing other performance indicators.

Self-assessment is closely tied to self-reflection-related concepts, such as self-observation, self-criticality, self-reaction, and self-concept. This research highlights the importance of self-assessment in triggering self-observation, a critical aspect of learning communication skills. Self-observation is necessary for accurate self-assessment as the evidence it provides makes the assessment more informed.

Given that learning communication skills is often implicit, self-assessment can help bring communication experiences into conscious awareness, a challenging yet essential process (Anders Ericsson, 2008; Eraut*, 2004; Zinn, 1988). Once students become aware of their communication behaviours, performance revision tends to occur naturally (Giroldi et al., 2017; Verheijden et al., 2023).

Interestingly, students with low patient assessments but accurate self-assessments demonstrated strong self-criticality, identifying areas for improvement more precisely than previous research suggests (Davis et al., 2006; Dunning, 2011). These students were honest with themselves, recalling their actions and making accurate judgments about their performance. Conversely, high-performing students with lower self-assessments often held themselves to higher standards, reflecting a more nuanced understanding of their performance. This is consistent with findings that high performers tend to be stringent self-assessors (Brown et al., 2015; Coutts & Rogers, 1999; Edwards et al., 2003). Maybe it is, as Charles Darwin once said, that "*..ignorance more frequently begets confidence than does knowledge*" (Darwin, 1874, p. 3).

As for students whose self-concept may have interfered with their self-assessment, Brown et al. (2015, p. 451) insightfully suggested: *"to help students focus on the quality of their work, as an independent artefact, rather than an extension of their ego.*" This highlights how self-assessment can be a misnomer since it is not an activity where someone

evaluates the self (what and who they are) (Kasanen & Räty, 2002). Therefore, researchers should try to remove ego or at least account for it in self-assessment (Brown et al., 2015).

5.4 CONTRIBUTION TO THE RELEVANT ACADEMIC LITERATURE

This study contributes significantly by demonstrating how self-assessment can be integrated into external assessment systems to enhance performance. The integrated findings show that students largely engage in accurate self-analysis, which might not reflect consistently between students due to the differential use and interpretation of the scale. A key contribution is uncovering the hidden meanings behind self-assessment scores, often overlooked in monomethod studies that lack a qualitative depth (Cevallos et al., 2024; Evans et al., 2007; León et al., 2023; Mays & Branch-Mays, 2016; Nayar et al., 2020; Sun et al., 2020).

The study shows that self-assessment positively influences communication skills development, a complex area to define and assess compared to technical skills (Makoul, 2001b; van den Eertwegh et al., 2015). This difficulty is multiplied when asking inexperienced assessors, such as students, to engage in meaningful self-assessment (Blanch-Hartigan, 2011). Thus, designing, validating, and demonstrating the educational impact of a communication self-assessment instrument is a significant addition to the literature.

Furthermore, this research introduces and validates the first Arabic communication assessment tool based on the Calgary-Cambridge Guide, crucial for addressing communication skills training gaps at Damascus University and adaptable for other Arabic dental and medical schools, where communication training is often underdeveloped (AL-Shehri, 2012; Bylund et al., 2017).

5.5 STRENGTHS AND LIMITATIONS

This study is among the first mixed-methods research on self-assessment of communication skills, combining quantitative and qualitative data from students' self-assessments and assessments by real and standardised patients. A comprehensive understanding of the assessment method's validity was provided, covering its

psychometric properties, educational impact, and the factors influencing self-assessment scoring. Integrating qualitative and quantitative findings allowed for a deeper understanding of how each strand complements or challenges the other, allowing broader and more corroborated conclusion to be drawn.

In the quantitative strand, the first generalizability study for self-assessment in direct observation of communication skills was conducted. However, a limitation is the conflation of person and rater variance, complicating the interpretation of the G-coefficient, an issue not addressed in Peterson et al. (2012) study. A limitation of the EFA, noted in the methods section (**see Section 3.5.4**), is its inability to address the multi-level structure of the data. Ideally, this requires multi-level confirmatory factor analysis (CFA). However, I did not perform this due to my unfamiliarity with multi-level CFA. Despite this, the split-half EFA yielded results similar to those from the full dataset, which, to a degree, support the reliability of the output.

A key strength of the qualitative strand was integrating quantitative findings into the interviews. This approach facilitated targeted participant selection, served as a focal point of discussion, reduced recall bias and enabled cross-validation between students' accounts and patient assessments. Although direct coding without transcription is debatable, it was necessary due to time, resource constraints, and a lack of accurate Arabic transcription software. Although transcription is typically viewed as essential, some scholars argue it is resource-intensive and prone to errors (Clausen, 2012; Halcomb & Davidson, 2006). Drawing from my own and other researchers' experiences, direct coding "*offers the researcher the opportunity to hear 'the music beyond the words' (i.e., the subtle pauses, volume fluctuations, interruptions, etc.)"* likely to be missed in most transcription approaches (Gravois et al., 1992, p. 567). Gravois et al. (1992) and Mayerovitch (1972) showed evidence that direct coding of audiotapes can be reliable. Nonetheless, this method was labour-intensive, requiring repeated listening to refine codes.

5.6 DIRECTIONS FOR FUTURE RESEARCH

This study faced limitations due to the practical constraints of being conducted by a single researcher within the timeframe of an MSc dissertation. Additionally, the study's scope was bounded by its specific aims and research questions. Future research could build on this study in several valuable ways:

1. Conduct a large-scale survey to test the qualitative insights on a broader sample.
2. Given the reliability observed in real patient scores, training patients to enhance their role as assessors could be beneficial.
3. Future studies could explore communication skills within dental treatment sessions and post-operatively, examining the interplay between biomedical, procedural, and communication skills.
4. Investigating self-assessment over an extended period would provide deeper insights into its long-term impact on skill development.
5. Similar research in diverse settings and cultures would be essential to validate the assessment approach across various contexts.

5.7 IMPLICATIONS FOR EDUCATIONAL PRACTICE

The introduced assessment approach is a valuable tool for assessment-for-learning in dental schools, particularly for communication skills training. It encourages students to self-assess and identify areas for improvement. However, seeking feedback from supervisors, peers, or patients is essential, as self-assessment alone may overlook performance gaps. While this approach provides insights, it does not guarantee skill improvement; the responsibility to act on these insights lies with the student.

To effectively use self-assessment data, students must receive training and feedback on accurate self-assessment. Research indicates that subjective use of assessment forms can be minimised with proper guidance (Alfakhry, Mustafa, Ybrode, et al., 2023; León et al., 2023). Faculty should make it clear to students that self-assessment communication performance is an independent artefact in a manner that keeps their ego intact and focuses on future-oriented actions (Brown et al., 2015). It is equally important not to tempt students to distort their self-assessment by keeping the purpose formative and educational (Brown et al., 2015). Lastly, faculty should not overly focus on how accurately students self-assess, as it is highly contextual and varies with the assessment method and skill being evaluated (Bol & Hacker, 2012; Eva & Regehr, 2008; Maki et al., 2005). Self-assessment should be understood as a means to an end, the end being the specific skill under evaluation.

REFERENCES

ACGME. (2023). *Accreditation Council for Graduate Medical*

Education Common Program Requirements (Residency). Retrieved 09/12/2023 from https://www.acgme.org/programs-and-institutions/programs/common-program-requirements/

Acheampong, D., Ludvik, M. B., & Lundquist, A. E. (2023). Developing Individual Awareness: The Role of the Assessor. In *Reframing Assessment to Center Equity* (pp. 303-325). Routledge.

AERA/APA/NCME. (1999). Standards for educational and psychological testing. *New York: American Educational Research Association*.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational research methods*, *16*(2), 270-301.

AL-Shehri, M. Y. (2012). Communication skills courses: a plea for inclusion in the medical curriculum. *Saudi Journal for Health Sciences*, *1*(1), 2-4.

Alfakhry, G., Mustafa, K., Alagha, M. A., Milly, H., Dashash, M., & Jamous, I. (2022). Bridging the gap between self-assessment and faculty assessment of clinical performance in restorative dentistry: A prospective pilot study. *Clinical and Experimental Dental Research*, *8*(4), 883-892.

Alfakhry, G., Mustafa, K., AlMukhallalati, A., Alhomsi, K., Saymeh, R., & Jamous, I. (2023). Evaluation of the Undergraduate Learning Environment at Dental Schools in Syria. *international dental journal*, *73*(5), 659-666. https://doi.org/https://doi.org/10.1016/j.identj.2022.12.001

Alfakhry, G., Mustafa, K., Jazayerli, B., Alhomsi, K., & Jamous, I. (2023). The challenging reality of the clinical learning environment at Damascus University Faculty of Dental Medicine in Syria: a qualitative study. *MedEdPublish*, *13*(24), 24.

Alfakhry, G., Mustafa, K., Ybrode, K., Jazayerli, B., Milly, H., Abohajar, S., Hassan, H., Alhomsi, K., & Jamous, I. (2023). Evaluation of a workplace assessment method designed to improve self-assessment in operative dentistry: a quasi-experiment. *BMC Medical Education*, *23*(1), 491.

Anders Ericsson, K. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Academic emergency medicine*, *15*(11), 988-994.

Andersen, S. A. W., Nayahangan, L. J., Park, Y. S., & Konge, L. (2021). Use of generalizability theory for exploring reliability of and sources of variance in assessment of technical skills: A systematic review and meta-analysis. *Academic Medicine*, *96*(11), 1609-1619.

Andrade, H. L., & Brown, G. T. (2016). Student self-assessment in the classroom. In *Handbook of human and social conditions in assessment* (pp. 319-334). Routledge.

Artino, A. R. (2012). Academic self-efficacy: from educational theory to instructional practice. *Perspectives on medical education*, *1*, 76-85.

Ashton, C. M., Haidet, P., Paterniti, D. A., Collins, T. C., Gordon, H. S., O'Malley, K., Petersen, L. A., Sharf, B. F., Suarez-Almazor, M. E., & Wray, N. P. (2003). Racial and ethnic disparities in the use of health services: bias, preferences, or poor communication? *Journal of general internal medicine*, *18*(2), 146-152.

Bandura, A., & Wessels, S. (1997). *Self-efficacy*. Cambridge University Press Cambridge.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (Vol. 3). Wiley New York.

Barrows, H. S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Academic Medicine*, *68*(6), 443-451.
Bazeley, P. (2024). Conceptualizing Integration in Mixed Methods Research. *Journal of Mixed Methods Research*, 15586898241253636.
Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186-3191.
Bensing, J., Van Dulmen, S., & Tates, K. (2003). Communication in context: new directions in communication research. *Patient Education and Counseling*, *50*(1), 27-32.
BERA. (2018). *Ethical guidelines for educational research*. Bera.
Bertakis, K. D. (2009). The influence of gender on the doctor–patient interaction. *Patient Education and Counseling*, *76*(3), 356-360.
Bird, F. (2020). A defense of objectivity in the social sciences, rightly understood. *Sustainability: Science, Practice and Policy*, *16*(1), 83-98.
Blanch-Hartigan, D. (2011). Medical students' self-assessment of performance: results from three meta-analyses. *Patient Education and Counseling*, *84*(1), 3-9.
Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Medical Teacher*, *34*(11), 960-992.
Bokken, L., Rethans, J.-J., van Heurn, L., Duvivier, R., Scherpbier, A., & van der Vleuten, C. (2009). Students' views on the use of real patients and simulated patients in undergraduate medical education. *Academic Medicine*, *84*(7), 958-963.
Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self assessment and reflection in professional courses. *Social work education*, *18*(2), 121-132.
Bowen, J. L., Ilgen, J. S., Regehr, G., Cate, O. t., Irby, D. M., & O'Brien, B. C. (2019). Reflections From the Rearview Mirror: Internal Medicine Physicians' Reactions to Clinical Feedback After Transitions of Responsibility. *Academic Medicine*, *94*(12), 1953-1960. https://doi.org/10.1097/acm.0000000000002831
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.
Brennan, R. L. (2001). *Generalizability Theory*. Springer-Verlag Publishing.
Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 444-457.
Brown, R. F., Wuensch, A., & Bylund, C. L. (2017). Models of communication skills training and their practical implications. *Oxford textbook of communication in oncology and palliative care*, 16-22.
Bylund, C. L., Scott, S., & Alyafei, K. (2017). Communication skills training in Arab countries: Opportunities and challenges in the Qatar experience. *Oxford textbook of communication in oncology and palliative care*, 392.
Carey, J. A., Madill, A., & Manogue, M. (2010). Communications skills in dental education: a systematic research review. *European Journal of Dental Education*, *14*(2), 69-78.
Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, *1*(2), 245-276.
Cevallos, M., Pardo, A., Kakadia, R., Jiang, T., & Chutinan, S. (2024). Can self-assessment ability predict student performance in a dental anatomy course? *European Journal of Dental Education*, *28*(1), 79-85. https://doi.org/https://doi.org/10.1111/eje.12916

Claramita, M., Nugraheni, M. D., van Dalen, J., & van der Vleuten, C. (2013). Doctor–patient communication in Southeast Asia: a different culture? *Advances in health sciences education*, *18*, 15-31.
Clausen, A. S. (2012). The individually focused interview: Methodological quality without transcription of audio recordings. *The Qualitative Report*, *17*(19), 1-17.
Cohen, J. (1988). *Statistical power analysis for the behavioural sciences (2nd edition). Hillsdale, NJ: Erlbaum.*
Cook-Sather, A., & Matthews, K. E. (2021). Pedagogical partnership:: Engaging with students as co-creators of curriculum, assessment and knowledge. In *University teaching in focus* (pp. 243-259). Routledge.
Cook, D. A., & Beckman, T. J. (2009). Does scale length matter? A comparison of nine-versus five-point rating scales for the mini-CEX. *Advances in health sciences education*, *14*(5), 655-664. https://doi.org/10.1007/s10459-008-9147-x
Côté, L., Savard, A., & Bertrand, R. (2001). [Evaluation of the physician-patient relationship competence. Development and validation of an assessment instrument]. *Can Fam Physician*, *47*, 512-518. (Evaluation des compétences relationnelles du médecin avec le patient. Elaboration et validation d'une grille.)
Coutts, L., & Rogers, J. (1999). Predictors of student self-assessment accuracy during a clinical performance exam: comparisons between over-estimators and under-estimators of SP-evaluated performance. *Academic Medicine*, *74*(10), S128-130. https://journals.lww.com/academicmedicine/fulltext/1999/10000/predictors_of_student_self_assessment_accuracy.62.aspx
Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research* (Third Edition. International Student Edition ed.). Sage publications.
Cushing, A. (1996). Communication skills. *Medical education*, *30*(5), 316-318. https://doi.org/https://doi.org/10.1111/j.1365-2923.1996.tb00839.x
Daniels, V. J., & Harley, D. (2017). The effect on reliability and sensitivity to level of training of combining analytic and holistic rating scales for assessing communication skills in an internal medicine resident OSCE. *Patient Education and Counseling*, *100*(7), 1382-1386. https://doi.org/https://doi.org/10.1016/j.pec.2017.02.014
Darwin, C. (1874). *The descent of man : and selection in relation to sex*. New York : A.L. Burt.
Dashash, M. (2019). Unified national medical assessment: an approach towards meeting the needs of the population during the Syrian Crisis. *Major Challenges Facing Higher Education in the Arab World: Quality Assurance and Relevance*, 363-382.
Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of Physician Self-assessment Compared With Observed Measures of CompetenceA Systematic Review. *Jama*, *296*(9), 1094-1102. https://doi.org/10.1001/jama.296.9.1094
De Vaus, D. (2001). Research design in social research. *Research design in social research*, 1-296.
de Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, *39*(4), 695-710. https://doi.org/10.1080/02664763.2011.610445
Dewey, J. ((1948, 1920)). *Reconstruction in philosophy*. Beacon Press.
Duffy, F. D., & Holmboe, E. S. (2006). Self-assessment in lifelong learning and improving performance in practice: physician know thyself. *Jama*, *296*(9), 1137-1139.

Dunning, D. (2011). Chapter five - The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 44, pp. 247-296). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-385522-0.00005-6

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological science in the public interest*, *5*(3), 69-106.

Edwards, R. K., Kellner, K. R., Sistrom, C. L., & Magyari, E. J. (2003). Medical student self-assessment of performance on an obstetrics and gynecology clerkship. *American Journal of Obstetrics and Gynecology*, *188*(4), 1078-1082. https://doi.org/https://doi.org/10.1067/mob.2003.249

El Gayar, N., & Suen, C. Y. (2018). *Computational linguistics, speech and image processing for Arabic language* (Vol. 4). World Scientific.

Ende, J. (1983). Feedback in clinical medical education. *Jama*, *250*(6), 777-781.

Epstein, R. M., Siegel, D. J., & Silberman, J. (2008). Self-monitoring in clinical practice: a challenge for medical educators. *Journal of Continuing Education in the Health Professions*, *28*(1), 5-13.

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British journal of educational psychology*, *70*(1), 113-136.

Eraut*, M. (2004). Informal learning in the workplace. *Studies in continuing education*, *26*(2), 247-273.

Eva, K. W., & Regehr, G. (2005). Self-Assessment in the Health Professions: A Reformulation and Research Agenda. *Academic Medicine*, *80*(10), S46-S54. https://journals.lww.com/academicmedicine/fulltext/2005/10001/self_assessment_in_the_health_professions__a.15.aspx

Eva, K. W., & Regehr, G. (2008). “I'll never play professional football” and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions*, *28*(1), 14-19.

Eva, K. W., & Regehr, G. (2011). Exploring the divergence between self-assessment and self-monitoring. *Advances in health sciences education*, *16*, 311-329.

Evans, A. W., Leeson, R. M. A., & Petrie, A. (2007). Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery. *Medical education*, *41*(9), 866-872. https://doi.org/https://doi.org/10.1111/j.1365-2923.2007.02819.x

Field, A. (2017). *Discovering statistics using IBM SPSS statistics (fifth edition)*. sage.

Fitzgerald, J. T., White, C. B., & Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy. *Medical education*, *37*(7), 645-649.

Fox, C. (2010). Evidence summary: what do we know from qualitative research about people's care-seeking about oral health? *British Dental Journal*, *209*(5), 225-231.

Frambach, J. M., van der Vleuten, C. P., & Durning, S. J. (2013). AM last page: Quality criteria in qualitative and quantitative research. *Academic Medicine*, *88*(4), 552.

Gärtner, J., Bußenius, L., Schick, K., Prediger, S., Kadmon, M., Berberat, P. O., & Harendza, S. (2022). Validation of the ComCare index for rater-based assessment of medical communication and interpersonal skills. *Patient Education and Counseling*, *105*(4), 1004-1008. https://doi.org/https://doi.org/10.1016/j.pec.2021.07.051

GDC. (2015). *Preparing for practice: Dental team learning outcomes for registration (2015 revised edition)*. Retrieved 09/12/2023 from https://www.gdc-uk.org/education-cpd/quality-assurance

Gillespie, A. (2018). 13 Distinguishing Two Processes of Self-reflection. *The Cambridge handbook of sociocultural psychology*, 245.
Giroldi, E., Veldhuijzen, W., Geelen, K., Muris, J., Bareman, F., Bueving, H., van der Weijden, T., & van der Vleuten, C. (2017). Developing skilled doctor–patient communication in the workplace: a qualitative study of the experiences of trainees and clinical supervisors. *Advances in health sciences education*, *22*, 1263-1278.
GMC. (2023). *Professional standards for doctors*. Retrieved 09/12/2023 from https://www.gmc-uk.org/professional-standards/professional-standards-for-doctors#professionalism
Gobat, N., Kinnersley, P., Gregory, J. W., Pickles, T., Hood, K., & Robling, M. (2015). Measuring clinical skills in agenda-mapping (EAGL-I). *Patient Education and Counseling*, *98*(10), 1214-1221. https://doi.org/https://doi.org/10.1016/j.pec.2015.06.018
Gorsuch, R. L. (1983). *Factor Analysis (2nd ed.)*. Psychology Press. https://doi.org/https://doi.org/10.4324/9780203781098
Govaerts, M., Van de Wiel, M., Schuwirth, L., Van der Vleuten, C., & Muijtjens, A. (2013). Workplace-based assessment: raters' performance theories and constructs. *Advances in health sciences education*, *18*, 375-396.
Govaerts, M., & van der Vleuten, C. P. (2013). Validity in work-based assessment: expanding our horizons. *Medical education*, *47*(12), 1164-1174.
Govaerts, M. J. B., van der Vleuten, C. P. M., Schuwirth, L. W. T., & Muijtjens, A. M. M. (2007). Broadening Perspectives on Clinical Performance Assessment: Rethinking the Nature of In-training Assessment. *Advances in health sciences education*, *12*(2), 239-260. https://doi.org/10.1007/s10459-006-9043-1
Gravois, T., Rosenfield, S., & Greenberg, B. (1992). Establishing Reliability for Coding Implementation Concerns of School-Based Teams from Audiotapes. *Evaluation Review*, *16*(5), 562-569. https://doi.org/10.1177/0193841x9201600507
Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis*, *11*(3), 255-274.
Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological bulletin*, *103*(2), 265.
Hager, P. (2011). Theories of workplace learning. *The SAGE handbook of workplace learning*, 17-31.
Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis. In *Multivariate data analysis* (pp. 785-785).
Halcomb, E. J., & Davidson, P. M. (2006). Is verbatim transcription of interview data always necessary? *Applied Nursing Research*, *19*(1), 38-42. https://doi.org/https://doi.org/10.1016/j.apnr.2005.06.001
Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological measurement*, *66*(3), 393-416.
Hodges, B., Turnbull, J., Cohen, R., Bienenstock, A., & Norman, G. (1996). Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. *Medical education*, *30*(1), 38-43. https://doi.org/https://doi.org/10.1111/j.1365-2923.1996.tb00715.x
Huebner, A., & Lucht, M. (2019). Generalizability theory in R. *Practical Assessment, Research, and Evaluation*, *24*(1), 5.
Israel, M., & Hay, I. (2006). *Research ethics for social scientists*. Sage.

Jetty, A., Jabbarpour, Y., Pollack, J., Huerto, R., Woo, S., & Petterson, S. (2022). Patient-Physician Racial Concordance Associated with Improved Healthcare Use and Lower Healthcare Expenditures in Minority Populations. *Journal of Racial and Ethnic Health Disparities*, *9*(1), 68-81. https://doi.org/10.1007/s40615-020-00930-4

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, *33*(7), 14-26.

Kaiser, H. F. (1970). A second generation little jiffy.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological measurement*, *34*(1), 111-117.

Karabilgin, O. S., Vatansever, K., Caliskan, S. A., & Durak, H. İ. (2012). Assessing medical student competency in communication in the pre-clinical phase: Objective structured video exam and SP exam. *Patient Education and Counseling*, *87*(3), 293-299. https://doi.org/https://doi.org/10.1016/j.pec.2011.10.008

Kasanen, K., & Räty, H. (2002). "You be Sure Now to be Honest in Your Assessment": Teaching and Learning Self-Assessment. *Social Psychology of Education*, *5*(4), 313-328. https://doi.org/10.1023/A:1020993427849

Keller, V. F., & Carroll, J. G. (1994). A new model for physician-patient communication. *Patient Education and Counseling*, *23*(2), 131-140.

Khalifah, A. M., & Celenza, A. (2019). Teaching and assessment of dentist-patient communication skills: a systematic review to identify best-evidence methods. *Journal of Dental Education*, *83*(1), 16-31.

Kruse, A. B., Heil, H. K. P., Struß, N., Fabry, G., Silbernagel, W., Vach, K., Ratka-Krüger, P., & Woelber, J. P. (2020). Working experience is not a predictor of good communication: Results from a controlled trial with simulated patients. *European Journal of Dental Education*, *24*(2), 177-185.

Kurtz, S., Draper, J., & Silverman, J. (1998). *Teaching and learning communication skills in medicine*.

Kurtz, S., Draper, J., & Silverman, J. (2017). *Teaching and learning communication skills in medicine*. CRC press.

Kurtz, S., Silverman, J., Benson, J., & Draper, J. (2003). Marrying content and process in clinical method teaching: enhancing the Calgary–Cambridge guides. *Academic Medicine*, *78*(8), 802-809.

Kurtz, S. M., & Silverman, J. D. (1996). The Calgary—Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes. *Medical education*, *30*(2), 83-89.

Lehdonvirta, V., Oksanen, A., Räsänen, P., & Blank, G. (2021). Social Media, Web, and Panel Surveys: Using Non-Probability Samples in Social and Policy Research. *Policy & Internet*, *13*(1), 134-155. https://doi.org/https://doi.org/10.1002/poi3.238

León, S. P., Panadero, E., & García-Martínez, I. (2023). How accurate are our students? A meta-analytic systematic review on self-assessment scoring accuracy. *Educational Psychology Review*, *35*(4), 106.

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. *The Sage handbook of qualitative research*, *4*(2), 97-128.

Linn, R. (1993). *Educational measurement* (Thrid Edition ed.).

Luong, V., Ajjawi, R., Burm, S., Olson, R., & MacLeod, A. (2024). Unravelling epistemic injustice in medical education: The case of the underperforming learner. *Medical education*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/medu.15410

Makoul, G. (2001a). Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Academic Medicine*, *76*(4), 390-393.
Makoul, G. (2001b). The SEGUE Framework for teaching and assessing communication skills. *Patient Education and Counseling*, *45*(1), 23-34.
Mann, K. V. (2011). Theoretical perspectives in medical education: past experience and future possibilities. *Medical education*, *45*(1), 60-68.
Maxcy, S. J. (2003). Pragmatic threads in mixed methods research in the social sciences: The search for multiple modes of inquiry and the end of the philosophy of formalism. *Handbook of mixed methods in social and behavioral research*(51-89).
Mayerovitch, J. (1972). *A RELIABILITY STUDY OF A NEW SYSTEM FOR CODING FAMILY THERAPISTS'VERBALIZATIONS*. Western Michigan University.
Mays, K. A., & Branch-Mays, G. L. (2016). A systematic review of the use of self-assessment in preclinical and clinical dental education. *Journal of Dental Education*, *80*(8), 902-913.
Monteiro, S., Sullivan, G. M., & Chan, T. M. (2019). Generalizability theory made simple (r): an introductory primer to G-studies. In (Vol. 11, pp. 365-370): The Accreditation Council for Graduate Medical Education.
Moore, C. T. (2016). *gtheory: Apply Generalizability Theory with R. R package version 0.1.2.* . Retrieved June 8, 2024 from https://CRAN.R-project.org/package=gtheory
Moorthy, K., Munz, Y., Adams, S., Pandey, V., Darzi, A., & Hospital, I. C. S. M. s. (2006). Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. *The American journal of surgery*, *192*(1), 114-118.
Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing research*, *40*(2), 120-123.
Moult, A., McKinley, R. K., & Yeates, P. (2021). Understanding patient involvement in judging students' communication skills in OSCEs. *Medical Teacher*, *43*(9), 1070-1078.
Nayar, S. K., Musto, L., Baruah, G., Fernandes, R., & Bharathan, R. (2020). Self-Assessment of Surgical Skills: A Systematic Review. *Journal of surgical education*, *77*(2), 348-361. https://doi.org/https://doi.org/10.1016/j.jsurg.2019.09.016
NHS. (2022). *Data on Written Complaints in the NHS, 2021-22 Quarter 3 and Quarter 4. NHS Digital.* Retrieved 09/12/2023 from https://digital.nhs.uk/data-and-information/publications/statistical/data-on-written-complaints-in-the-nhs/2021-22-quarter-3-and-quarter-4#
Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, *35*(4), 250.
Norman, G. (2005). Editorial–checklists vs. ratings, the illusion of objectivity, the demise of skills and the debasement of evidence. *Advances in health sciences education*, *10*(1), 1-3.
O'Hara, J. K., Reynolds, C., Moore, S., Armitage, G., Sheard, L., Marsh, C., Watt, I., Wright, J., & Lawton, R. (2018). What can patients tell us about the quality and safety of hospital care? Findings from a UK multicentre survey study. *BMJ quality & safety*, *27*(9), 673-682.
Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, *13*(1), 48-63.

Paasche-Orlow, M. K., Parker, R. M., Gazmararian, J. A., Nielsen-Bohlman, L. T., & Rudd, R. R. (2005). The prevalence of limited health literacy. *Journal of general internal medicine*, *20*(2), 175-184.

Paravattil, B., & Wilby, K. J. (2019). Optimizing assessors' mental workload in rater-based assessment: a critical narrative review. *Perspectives on medical education*, *8*(6), 339-345.

Patel, U. S., Tonni, I., Gadbury-Amyot, C., Van der Vleuten, C., & Escudier, M. (2018). Assessment in a global context: An international perspective on dental education. *European Journal of Dental Education*, *22*, 21-27.

Perera, J., Mohamadou, G., & Kaur, S. (2010). The use of objective structured self-assessment and peer-feedback (OSSP) for learning communication skills: evaluation using a controlled trial. *Advances in health sciences education*, *15*, 185-193.

Peterson, L. N., Eva, K. W., Rusticus, S. A., & Lovato, C. Y. (2012). The readiness for clerkship survey: can self-assessment data be used to evaluate program effectiveness? *Academic Medicine*, *87*(10), 1355-1360.

Phillips, D. C. (2014). Research in the hard sciences, and in very hard "softer" domains. *Educational researcher*, *43*(1), 9-11.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, *104*(1), 1-15.

Pring, R. (2000). The 'False Dualism' of Educational Research. *Journal of philosophy of education*, *34*(2), 247-260. https://doi.org/https://doi.org/10.1111/1467-9752.00171

Rethans, J. J., Gorter, S., Bokken, L., & Morrison, L. (2007). Unannounced standardised patients in real practice: a systematic literature review. *Medical education*, *41*(6), 537-549.

Rhodes, K. V., & Miller, F. G. (2012). Simulated patient studies: an ethical analysis. *The milbank quarterly*, *90*(4), 706-724.

Röttele, N., Schlett, C., Körner, M., Farin-Glattacker, E., Schöpf-Lazzarino, A. C., Voigt-Radloff, S., & Wirtz, M. A. (2021). Variance components of ratings of physician-patient communication: A generalizability theory analysis. *Plos one*, *16*(6), e0252968.

Salim, N. A., Sallam, M., Aldweik, R. e. H., Sawair, F. A., Sharaireh, A. M., & Alabed, A. (2023). Rating communication skills in dental practice: the impact of different sociodemographic factors. *BMC Medical Education*, *23*(1), 950.

Sambell, K. (2013). Involving students in the scholarship of assessment: Student voices on the feedback agenda for change. In *Reconceptualising Feedback in Higher Education* (pp. 80-91). Routledge.

Sargeant, J., Armson, H., Chesluk, B., Dornan, T., Eva, K., Holmboe, E., Lockyer, J., Loney, E., Mann, K., & van der Vleuten, C. (2010). The processes and dimensions of informed self-assessment: a conceptual model. *Academic Medicine*, *85*(7), 1212-1220.

Schaepkens, S. P. C., de la Croix, A., & Veen, M. (2024). 'Oh yes, that is also reflection'—Using discursive psychology to describe how GP registrars construct reflection. *Medical education*, *58*(3), 318-326. https://doi.org/https://doi.org/10.1111/medu.15183

Schönrock-Adema, J., Heijne-Penninga, M., Van Hell, E. A., & Cohen-Schotanus, J. (2009). Necessary steps in factor analysis: enhancing validation studies of

educational instruments. The PHEEM applied to clerks as an example. *Medical Teacher*, *31*(6), e226-e232.
Schönwetter, D. J., Wener, M. E., & Mazurat, N. (2012). Determining the validity and reliability of clinical communication assessment tools for dental patients and students. *Journal of Dental Education*, *76*(10), 1276-1290.
Schumacker, R. E. (2014). *Learning statistics using R*. Sage Publications.
Schuwirth, L. W., & Van der Vleuten, C. P. (2018). How to design a useful test: the principles of assessment. *Understanding medical education: evidence, theory, and practice*, 275-289.
Shariff, F., Hatala, R., & Regehr, G. (2020). Learning After the Simulation Is Over: The Role of Simulation in Supporting Ongoing Self-Regulated Learning in Practice. *Academic Medicine*, *95*(4), 523-526. https://doi.org/10.1097/acm.0000000000003078
Shehada, M. R., Alfakhry, G., Jamous, I., Aljoujou, A. A., & Abdul_hak, M. (2023). Major Stress Sources Amongst Dental Students at Damascus University, Syria. *international dental journal*, *73*(2), 205-211.
Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Zhong nan da xue xue bao. Yi xue ban= Journal of Central South University. Medical sciences*, *37*(2), 152-155.
Shibli-Rahhal, A., & Kreiter, C. (2021). Development and evaluation of novel tool to assess communication skills in adult triadic interviews. *Patient Education and Counseling*, *104*(10), 2412-2417. https://doi.org/https://doi.org/10.1016/j.pec.2021.07.001
Shigli, K., & Awinashe, V. (2010). Patient-dentist communication: an adjunct to successful complete denture treatment. *Journal of Prosthodontics: Implant, Esthetic and Reconstructive Dentistry*, *19*(6), 491-493.
Shrout, P. E., & Yager, T. J. (1989). Reliability and validity of screening scales: Effect of reducing scale length. *Journal of Clinical Epidemiology*, *42*(1), 69-78. https://doi.org/https://doi.org/10.1016/0895-4356(89)90027-9
Silverman, J., Archer, J., Gillard, S., Howells, R., & Benson, J. (2011). Initial evaluation of EPSCALE, a rating scale that assesses the process of explanation and planning in the medical interview. *Patient Education and Counseling*, *82*(1), 89-93. https://doi.org/https://doi.org/10.1016/j.pec.2010.02.022
Silverman, J., Kurtz, S., & Draper, J. (2016). *Skills for communicating with patients*. crc press.
Simek-Downing, L., Quirk, M., & Letendre, A. (1986). Simulated versus actual patients in teaching medical interviewing. *Family medicine*, *18*(6), 358-360.
Simmenroth-Nayda, A., Heinemann, S., Nolte, C., Fischer, T., & Himmel, W. (2014). Psychometric properties of the Calgary Cambridge guides to assess communication skills of undergraduate medical students. *International journal of medical education*, *5*, 212.
Solchanyk, D., Ekeh, O., Saffran, L., Burnett-Zeigler, I. E., & Doobay-Persaud, A. (2021). Integrating cultural humility into the medical education curriculum: strategies for educators. *Teaching and Learning in Medicine*, *33*(5), 554-560.
Sousa, V. D., Hartman, S. W., Miller, E. H., & Carroll, M. A. (2009). New measures of diabetes self-care agency, diabetes self-efficacy, and diabetes self-management for insulin-treated individuals with type 2 diabetes. *Journal of clinical nursing*, *18*(9), 1305-1312.

Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of evaluation in clinical practice*, *17*(2), 268-274.
Spoto-Cannons, A. C., Isom, D. M., Feldman, M., Zwygart, K. K., Mhaskar, R., & Greenberg, M. R. (2019). Differences in medical student self-evaluations of clinical and professional skills. *Advances in medical education and practice*, 835-840.
Srikumaran, D., Tian, J., Ramulu, P., Boland, M. V., Woreta, F., Wang, K. M., & Mahoney, N. (2019). Ability of ophthalmology residents to self-assess their performance through established milestones. *Journal of surgical education*, *76*(4), 1076-1087.
Stalmeijer, R. E., Dolmans, D. H., Wolfhagen, I. H., Peters, W. G., van Coppenolle, L., & Scherpbier, A. J. (2010). Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Advances in health sciences education*, *15*, 315-328.
Stobart, G. (2009). Determining validity in national curriculum assessments. *Educational Research*, *51*(2), 161-179.
Sun, J., Adegbosin, A. E., Reher, V., Rehbein, G., & Evans, J. (2020). Validity and reliability of a self-assessment scale for Dental and Oral Health student's perception of transferable skills in Australia. *European Journal of Dental Education*, *24*(1), 42-52. https://doi.org/https://doi.org/10.1111/eje.12466
Symons, A. B., Swanson, A., McGuigan, D., Orrange, S., & Akl, E. A. (2009). A tool for self-assessment of communication skills and professionalism in residents. *BMC Medical Education*, *9*, 1-7.
Tashakkori, A., & Teddlie, C. (2010). *Sage handbook of mixed methods in social & behavioral research* (2nd ed.). SAGE Publications, Inc.
Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. *The SAGE handbook of qualitative research in psychology*, *2*(17-37), 25.
Tucker, L., & MacCallum, R. (1997). *Exploratory Factor Analysis*.
van den Eertwegh, V., van der Vleuten, C., Stalmeijer, R., van Dalen, J., Scherpbier, A., & van Dulmen, S. (2015). Exploring residents' communication learning process in the workplace: a five-phase model. *Plos one*, *10*(5), e0125958.
van der Vleuten, C., van den Eertwegh, V., & Giroldi, E. (2019). Assessment of communication skills. *Patient Education and Counseling*, *102*(11), 2110-2113. https://doi.org/https://doi.org/10.1016/j.pec.2019.07.007
Velicer, W. F., & Fava, J. L. (1998). Affects of variable and subject sampling on factor pattern recovery. *Psychological methods*, *3*(2), 231.
Verheijden, M., Giroldi, E., van den Eertwegh, V., Luijkx, M., van der Weijden, T., de Bruin, A., & Timmerman, A. (2023). Identifying characteristics of a skilled communicator in the clinical encounter. *Medical education*, *57*(5), 418-429.
Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring Self-assessment: Current State of the Art. *Advances in health sciences education*, *7*(1), 63-80. https://doi.org/10.1023/A:1014585522084
Wass, V., Van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *The Lancet*, *357*(9260), 945-949.
Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219-246. https://doi.org/10.1177/0095798418771807
WFME. (2015). *WFME Standards For Basic Medical Education*. Retrieved 09/12/2023 from https://wfme.org/download/wfme-global-standards-for-quality-improvement-2015-bme/

White, C. B., Gruppen, L. D., & Fantone, J. C. (2013). Self-regulated learning in medical education. In *Understanding Medical Education* (pp. 201-211). https://doi.org/https://doi.org/10.1002/9781118472361.ch15

Wilkinson, J. R., Crossley, J. G., Wragg, A., Mills, P., Cowan, G., & Wade, W. (2008). Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Medical education*, *42*(4), 364-373.

Williams, G. C., Saizow, R. B., & Ryan, R. M. (1999). The importance of self-determination theory for medical education. *Academic Medicine*, *74*(9), 992-995.

Williams, R. (2022). *The assessment of surgeon-patient communication: An examination of validity* University of Oxford].

Yan, Z. (2018). Student self-assessment practices: The role of gender, school level and goal orientation. *Assessment in Education: Principles, Policy & Practice*, *25*(2), 183-199.

Yates, N., Gough, S., & Brazil, V. (2022). Self-assessment: With all its limitations, why are we still measuring and teaching it? Lessons from a scoping review. *Medical Teacher*, *44*(11), 1296-1302. https://doi.org/10.1080/0142159X.2022.2093704

Yon, E. (2020). OSCEs: do you know how? *Veterinary Nursing Journal*, *35*(9-12), 269-271.

Yudkowsky, R., Downing, S. M., & Sandlow, L. J. (2006). Developing an institution-based assessment of resident communication and interpersonal skills. *Academic Medicine*, *81*(12), 1115-1122.

Zanetti, M., Keller, L., Mazor, K., Carlin, M., Alper, E., Hatem, D., Gammon, W., & Pugnaire, M. (2010). Using standardized patients to assess professionalism: a generalizability study. *Teaching and Learning in Medicine*, *22*(4), 274-279.

Zijlstra, W. P., van der Ark, L. A., & Sijtsma, K. (2011). Outliers in questionnaire data: Can they be detected and should they be removed? *Journal of Educational and Behavioral Statistics*, *36*(2), 186-212.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13-39). Elsevier.

Zinn, W. M. (1988). Doctors have feelings too. *Jama*, *259*(22), 3296-3298.

APPENDICES

APPENDIX I: First version of the developed assessment instrument

Table 16. The first version of the developed assessment instrument.

	Item
1	Greets you and introduces self appropriately
2	Identifies the reason for your visit accurately
3	Clearly outlines what will occur during the session
4	Employs effective questioning techniques, (using both open and closed questions and justifies the use of seemingly unrelated questions where appropriate)
5	Explores your thoughts, concerns, and emotional state regarding the problem
6	Requests consent for examination
7	Explains the reasons and steps involved in the clinical examination with respect to the level of information you want (brief, elaborate)
8	Handles interactions with sensitivity and understanding
9	Proactively informs the patient about potential discomfort during the examination
10	Delivering the information at appropriate level of understanding
11	Acknowledges deficiency in knowledge or skill (Requests help/time to consult someone when necessary)
12	Discusses treatment options, including costs, benefits, and risks, offering choice as applicable
13	Motivates the patient to adhere to the agreed-upon treatment plan (gives ideas of motivation, encouraging benefits, awareness of culture, empowers the patient)
14	Summarising key points in the session
15	Provides a clear plan for the next session
16	Appropriately address any remaining questions, concerns, or requests (relaying any remaining questions to the next session if time does not allow)
17	Providing safety nets (what to do in case things went in an unpleasant direction)
18	Treating you with respect
19	Demonstrates active listening
20	Demonstrates genuine care and concern for you as a person and responds with empathy
21	Displays humility in interacting with you
22	Uses appropriate verbal and nonverbal communication effectively
23	Inspires trust/reassurance
24	Organizes the session well (time, sequence of events, systematic, signposting)
25	Personalizes the session to you (is not robotic).
26	Has a professional appearance
27	Ensures that both of you have a shared understanding throughout the session
28	Encourages involvement and shared decision making in the session according with your preference

APPENDIX II: Pilot testing participants' characteristics and findings.

Table 17. Participants' characteristics for the pilot test based on self-report.

	Patients	student dentists	Experts
Number of participants	18	17	2
Age (mean±SD)	30.17±14.5	22.7±2.0	43
Sex (female %)[a]	82.30%	26.60%	Male
Education level			
9th grade	6	0	0
Baccalaureate[b]	3	0	0
Undergraduate	5	13	0
Postgraduate	0	4	0
PhD	0	0	2

a. Given the cultural context, sex was binary b. Baccalaureate is same as A-levels in the UK.

Table 18. Pilot testing results of the translated instrument by patients, clinical dental students and experts.

	Item no.	Patients (n=18)		Clinical dental students (n=17)		Experts (n=2)	
		Clarity	Relevance	Clarity	Relevance	Clarity	Relevance
1	Greets you and introduces self appropriately	94%	100%	100%	100%	100%	100%
2	Identifies the reason for your visit accurately	88%	75%	100%	88%	100%	100%
3	Clearly outlines what will occur during the session	100%	93%	94%	87%	100%	100%
4	Employs effective questioning techniques, (using both open and closed questions and justifies the use of seemingly unrelated questions where appropriate)	88%	94%	100%	100%	100%	100%
5	Explores your thoughts, concerns, and emotional state regarding the problem	88%	80%	94%	50%	100%	100%
6	Requests consent for examination	94%	93%	100%	63%	100%	100%
7	Explains the reasons and steps involved in the clinical examination with respect to the level of information you want (brief, elaborate)	100%	93%	94%	93%	100%	100%
8	Handles interactions with sensitivity and understanding	100%	93%	94%	88%	100%	100%
9	Proactively informs the patient about potential discomfort during the examination	89%	87%	100%	93%	100%	100%
10	Delivering information about diagnosis, causes, mechanism of diseases in a simple manner	94%	100%	100%	100%	100%	100%
11	Acknowledges deficiency in knowledge or skill (Requests help/time to consult someone when necessary)	100%	100%	94%	69%	100%	100%
12	Discusses treatment options, including costs, benefits, and risks, offering choice as applicable	94%	94%	100%	88%	100%	100%
13	Motivates the patient to adhere to the agreed-upon treatment plan (gives ideas of motivation, encouraging benefits, awareness of culture, empowers the patient)	100%	100%	94%	94%	100%	100%
14	Summarising key points in the session	94%	75%	88%	73%	100%	50%
15	Provides a clear plan for the next session	94%	88%	94%	63%	100%	100%
16	Appropriately address any remaining questions, concerns, or requests (relaying any remaining questions to the next session if time does not allow)	94%	100%	100%	75%	100%	100%
17	Providing safety nets (what to do in case things went in an unpleasant direction)	83%	88%	94%	87%	100%	100%
18	Treating you with respect	100%	100%	100%	94%	100%	100%
19	Demonstrates active listening	100%	100%	100%	100%	100%	100%
20	Demonstrates genuine care and concern for you as a person and responds with empathy	94%	88%	100%	63%	100%	100%
21	Displays humility in interacting with you	100%	100%	100%	88%	100%	100%
22	Uses appropriate verbal and nonverbal communication effectively	94%	89%	82%	81%	100%	100%
23	Inspires trust/reassurance	100%	94%	100%	94%	100%	100%
24	Organizes the session well (time, sequence of events, systematic, signposting)	94%	83%	88%	69%	100%	100%
25	Personalizes the session to you (is not robotic).	100%	88%	94%	75%	100%	100%
26	Has a professional appearance	94%	71%	100%	63%	100%	100%
27	Ensures that both of you have a shared understanding throughout the session	100%	100%	94%	81%	100%	100%
28	Encourages involvement and shared decision making in the session according with your preference	100%	94%	100%	69%	100%	100%

The percentages represent those who marked the item “clear” or “relevant”. Red-coloured values are less than 80%.

APPENDIX III: Assessment instrument (English and Arabic versions)

Communication Skills Assessment Form (self-assessment form)

Student code:…… Patient code (filled by the patient) :……

Date: …./…./2024 Duration of session: ……(20 minutes max)

Patient order: First/ second/ third/ fouth/ fifth/ sixth

Kindly assess your communication skills on the given 1-9 scale, put a circle on the number that reflects your assessment. The higher the number, the better the assessment. Kindly leave any comments that explain your ratings for each section

	Needs improvement			**Satisfactory**			**Excellent**		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Initiating the session and establishing rapport									
1. Greetings and introducing yourself to the patient appropriately	1	2	3	4	5	6	7	8	9
2. Identifying the reason for the visit accurately	1	2	3	4	5	6	7	8	9
3. Clearly outlining what will occur during the session	1	2	3	4	5	6	7	8	9
Important: Please share any comments that explain your ratings in the previous section:									
Exploring the problem with the patient									
4. Employing effective questioning techniques, (using both open and closed questions and justified the use of seemingly unrelated questions where appropriate)	1	2	3	4	5	6	7	8	9
5. Exploring patients' thoughts, concerns, and emotional state regarding the problem	1	2	3	4	5	6	7	8	9
Important: Please share any comments that explain your ratings in the previous section:									
Physical and Oral Examination									
6. Requesting consent	1	2	3	4	5	6	7	8	9
7. Explaining the reasons and steps involved in the clinical examination with respect to the level of information appropriate for the patient (brief, elaborate)	1	2	3	4	5	6	7	8	9
8. Handling interactions with sensitivity and understanding	1	2	3	4	5	6	7	8	9
9. Proactively informing the patient about potential discomfort during the examination	1	2	3	4	5	6	7	8	9
Important: Please share any comments that explain your ratings in the previous section:									

Move to the next page

	Needs improvement			Satisfactory			Excellent		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Explanation, Planning and demonstration of knowledge									
10. Delivering information about diagnosis, causes, mechanism of diseases in a simple manner	1	2	3	4	5	6	7	8	9
11. Acknowledging deficiency in knowledge or skill (Requests help/time to consult someone when necessary)	1	2	3	4	5	6	7	8	9
12. Discussing treatment options, including costs, benefits, and risks, offering choice as applicable	1	2	3	4	5	6	7	8	9
13. providing preventive and treatment instructions	1	2	3	4	5	6	7	8	9
14. Motivating the patient to adhere to the agreed-upon treatment plan (giving motivating reasons, encouraging benefits, awareness of culture, empowers the patient)	1	2	3	4	5	6	7	8	9
Important: Please share any comments that explain your ratings in the previous section:									
Closing the session									
15. Summarising key points in the session	1	2	3	4	5	6	7	8	9
16. Providing a clear plan for the next session	1	2	3	4	5	6	7	8	9
17. Appropriately addressing any remaining questions, concerns, or requests (relaying any remaining questions to the next session if time does not allow)	1	2	3	4	5	6	7	8	9
18. Providing safety nets (what to do in case things went in an unpleasant direction)	1	2	3	4	5	6	7	8	9
Important: Please share any comments that explain your ratings in the previous section:									
Cross-stage items (demonstrated throughout the session)									
19. Treating the patient with respect	1	2	3	4	5	6	7	8	9
20. Demonstrating active listening	1	2	3	4	5	6	7	8	9
21. Demonstrating genuine care and concern for you as a person and responds with empathy	1	2	3	4	5	6	7	8	9
22. Encouraging the patient to ask questions about anything.									

	Needs improvement			Satisfactory			Excellent		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
23. Displaying humility in interacting with the patient	1	2	3	4	5	6	7	8	9
24. Using appropriate verbal and nonverbal communication effectively	1	2	3	4	5	6	7	8	9
25. Inspiring trust/reassurance	1	2	3	4	5	6	7	8	9
26. Organizing the interview well (time, sequence of events, systematic, signposting)	1	2	3	4	5	6	7	8	9
27. Personalizing the session to the patient (not robotic).	1	2	3	4	5	6	7	8	9
28. Have a professional appearance	1	2	3	4	5	6	7	8	9
29. Ensuring shared understanding throughout the session	1	2	3	4	5	6	7	8	9
30. Encouraging patient's involvement and shared decision making according with patients' preference	1	2	3	4	5	6	7	8	9
Important: Please share any comments that explain your ratings in the previous section:									
Overall judgement of your own communication skills	1	2	3	4	5	6	7	8	9
Very important: Please justify your overall judgement and give examples:									

Do you have any comments regarding your communication skills that were not mentioned in the assessment form?

.....

.....

.....

.....

Move to the next page

Kindly tick (✓) **the one main area of improvement and one area of strength** you have:

	Patient-centred care	**Effective communication**	**Respect and Empathy**	**Professionalism and organization**
Area of improvement				
Area of strength				

End of the Form

Note: self-assessment (for students) and patient assessment forms differ only in pronoun use: (first-person for students, third-person for patients).

Standardised patients had additional extra questions about diagnosis and treatment planning accuracy and they were:

Was the diagnosis correct? Yes / No

Was the treatment plan appropriate? Yes / No

استمارة تقييم مهارات التواصل (للطلاب)

رمز الطالب:..... رمز المريض (يكتب من المريض):.....

التاريخ:...../5./2024 مدة الجلسة:.....دقيقة (لا تتجاوز 20 دقيقة)

ترتيب المريض: الأول \ الثاني \ الثالث \ الرابع \ الخامس \ السادس

يرجى تقييم مهارات التواصل لذاتك مع المريض على مقياس من 1 إلى 9 للبنود التالية, ضع دائرة على الرقم الذي يعكس تقييمك. اختيارك لرقم أعلى يعكس تقييم **أفضل**. لطفا اترك أية ملاحظات تفسر تقييمك لكل قسم.

البند	يحتاج الى تحسين			مقبول			ممتاز		
	1	2	3	4	5	6	7	8	9
بدء الجلسة وبناء العلاقة									
1. رحبت بالمريض وقدمت نفسك بشكل مناسب	1	2	3	4	5	6	7	8	9
2. استقصيت سبب الزيارة بدقة	1	2	3	4	5	6	7	8	9
3. ذكرت بوضوح ما سيحدث خلال الجلسة	1	2	3	4	5	6	7	8	9
(هام): يرجى ذكر أية ملاحظات توضح تقييماتك في القسم السابق:									
استقصاء المشكلة مع المريض									
4. أحسنت طرح الأسئلة (مستعملا الأسئلة المفتوحة والمغلقة وعند الحاجة مبررا أسئلتك التي تبدو غير ذات صلة بمشكلتك)	1	2	3	4	5	6	7	8	9
5. استقصيت أفكار ومخاوف وحالة المريض العاطفية المتعلقة بالمشكلة.	1	2	3	4	5	6	7	8	9
(هام): يرجى ذكر أية ملاحظات توضح تقييماتك في القسم السابق:									
الفحص السريري والفموي									
6. طلبت الإذن عند إجراء الفحص	1	2	3	4	5	6	7	8	9
7. شرحت أسباب وخطوات الفحص السريري مراعيا حجم المعلومات الذي يناسب المريض (مختصر أو مفصل)	1	2	3	4	5	6	7	8	9
8. تعاملت مع المريض بانتباه وتفهم	1	2	3	4	5	6	7	8	9
9. تبلغ المريض بشكل استباقي عن الإنزعاج المحتمل أثناء الفحص	1	2	3	4	5	6	7	8	9
(هام): يرجى ذكر أية ملاحظات توضح تقييماتك في القسم السابق:									

انتقل للوجه الثاني

	يحتاج الى تحسين			مقبول			ممتاز		
	1	2	3	4	5	6	7	8	9
الشرح والتخطيط ومناقشة خيارات المعالجة									
10. إيصال المعلومات المتعلقة بالتشخيص وأسباب وآلية المشكلة بطريقة سهلة الفهم	1	2	3	4	5	6	7	8	9
11. تعترف بالنقص في المعرفة أو المهارة (تطلب المساعدة/الوقت لاستشارة شخص ما عند الضرورة)	1	2	3	4	5	6	7	8	9
12. مناقشة خيارات المعالجة من حيث التكلفة و الفوائد والمخاطر ، مراعيا رغبة المريض	1	2	3	4	5	6	7	8	9
13. تعطي تعليمات وقائية وعلاجية	1	2	3	4	5	6	7	8	9
14. تحفز المريض على الالتزام بخطة العلاج المتفق عليها (تعطي أفكارا محفزة، تؤكد على الفوائد، تبرز وعي لثقافة المريض، تعزز دور المريض)	1	2	3	4	5	6	7	8	9
(هام): يرجى ذكر أية ملاحظات توضح تقييماتك في القسم السابق:									
اختتام الجلسة									
15. تلخص النقاط الرئيسية في الجلسة	1	2	3	4	5	6	7	8	9
16. تقدم خطة واضحة للجلسة القادمة	1	2	3	4	5	6	7	8	9
17. تجيب على أي أسئلة أو مخاوف أو طلبات متبقية بشكل مناسب (مع ترك أي أسئلة متبقية إلى الجلسة التالية في حال ضيق الوقت)	1	2	3	4	5	6	7	8	9
18. تقوم بتوفير خطة واضحة لما يجب فعله في حال لم تسر الأمور بالشكل المطلوب	1	2	3	4	5	6	7	8	9
(هام): يرجى ذكر أية ملاحظات توضح تقييماتك في القسم السابق:									
البنود المشتركة بين المراحل (الموضحة خلال طيلة الجلسة)									
19. تعامل المريض باحترام	1	2	3	4	5	6	7	8	9
20. تستمع باهتمام	1	2	3	4	5	6	7	8	9

	يحتاج الى تحسين			مقبول			ممتاز		
	1	2	3	4	5	6	7	8	9
21. تبدي عناية صادقة واهتمام بالمريض كشخص وتتجاوب بتعاطف	1	2	3	4	5	6	7	8	9
22. تشجع المريض على الاستفسار عن أي امر	1	2	3	4	5	6	7	8	9
23. تتعامل مع المريض بتواضع	1	2	3	4	5	6	7	8	9
24. تستخدم التواصل اللفظي والجسدي بمهارة	1	2	3	4	5	6	7	8	9
25. تبعث الثقة والإطمئنان للمريض	1	2	3	4	5	6	7	8	9
26. تنظم الجلسة بشكل جيد (الوقت، تسلسل الأحداث، منهجي، يعنون المراحل)	1	2	3	4	5	6	7	8	9
27. تخصص الجلسة بما يتلائم مع المريض(لا تتعامل مع المريض بشكل روبوتي)	1	2	3	4	5	6	7	8	9
28. ذو مظهر لبق ومهني	1	2	3	4	5	6	7	8	9
29. يحرص على وجود فهم مشترك للأمور خلال الجلسة	1	2	3	4	5	6	7	8	9
30. تشجع المريض على المشاركة واتخاذ القرار معك بما يتوافق مع رغبة المريض	1	2	3	4	5	6	7	8	9
يرجى ذكر أية ملاحظات توضح تقييماتك في القسم السابق									
تقييمك العام لمهارات التواصل الخاصة بك 31.	1	2	3	4	5	6	7	8	9
<u>:(هام جدا): يرجى تبرير تقييمك العام مع إعطاء أمثلة</u>									

هل لديك أي ملاحظات تتعلق بمهارات التواصل الخاصة بك والتي لم يتم ذكرها في بنود استمارة التقييم؟

.....

.....

.....

انتقل للوجه الثاني

بعد مراجعتك للجدول المرفق مع الاستمارة, يرجى وضع إشارة (✓) على نقطة رئيسية (**واحدة فقط**) تتطلب التحسين و نقطة قوة رئيسية (**واحدة فقط**) لدى المتدرب.

	الرعاية المتمحورة حول المريض	**التواصل الفعال**	**الاحترام والتعاطف**	**المهنية والتنظيم**
نقطة تتطلب التحسين				
نقطة قوة				

نهاية الاستمارة

Appendix IV: G*Power configuration and output

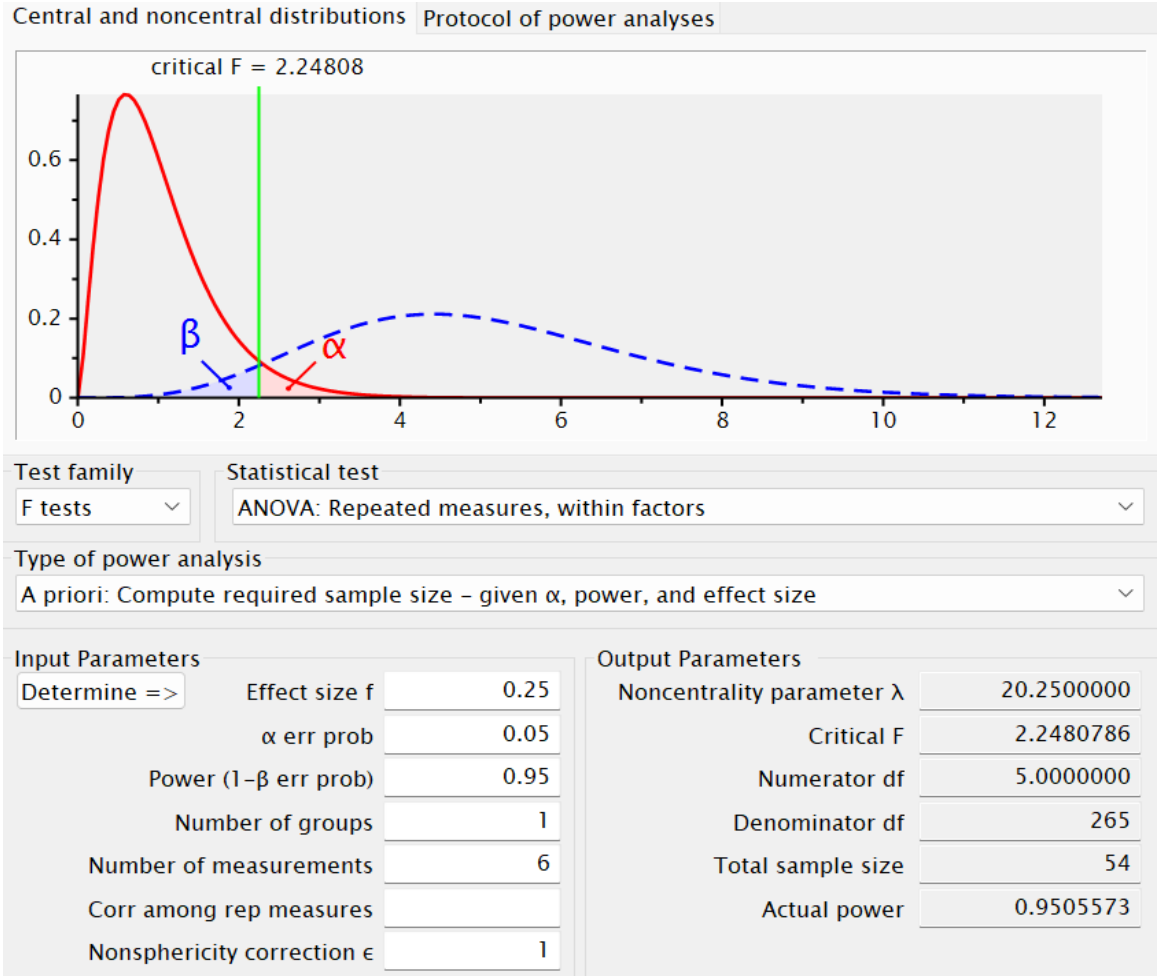


Figure 12 G*Power configuration to calculate the sample size

APPENDIX V: Missing data and outliers

Table 19. Item-non-response analysis for both self-assessment and patient assessment data.

Item no.	**Missing Self-assessment data**		**Missing Patient assessment data**	
	Number of missing data points	Percent	Number of missing data points	Percent
1	1	0.3%	0	0%
2	2	0.6%	0	0%
3	5	1.5%	0	0%
4	1	0.3%	0	0%
5	1	0.3%	0	0%
6	1	0.3%	0	0%
7	6	1.9%	2	0.6%
8	1	0.3%	5	1.5%
9	8	2.5%	0	0%
10	2	0.6%	0	0%
11	5	1.5%	1	0.3%
12	6	1.9%	0	0%
13	4	1.2%	1	0.3%
14	7	2.2%	4	1.2%
15	3	0.9%	0	0%
16	6	1.9%	1	0.3%
17	7	2.2%	0	0%
18	5	1.5%	0	0%
19	1	0.3%	1	0.3%
20	18	5.6%	1	0.3%
21	1	0.3%	8	2.5%
22	1	0.3%	0	0%
23	6	1.9%	2	0.6%
24	4	1.2%	1	0.3%
25	6	1.9%	7	2.2%
26	4	1.2%	5	1.5%
27	3	0.9%	0	0%
28	11	3.4%	2	0.6%
29	10	3.1%	3	0.9%
30	2	0.6%	0	0%
31	7	2.2%	0	0%

Table 20. Mean and standard deviation before and after deleting outliers for the self-assessment and patient assessment dataset.

	Self-assessment				Patient assessment			
Item no.	With outliers		Without outliers		With outliers		Without outliers	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	7.15	1.44	7.24	1.34	7.27	1.55	7.54	1.26
2	7.56	1.20	7.65	1.05	7.70	1.17	7.90	0.94
3	6.85	1.59	6.98	1.45	6.88	1.55	7.15	1.31
4	6.79	1.45	6.95	1.22	6.97	1.46	7.26	1.17
5	6.55	1.60	6.71	1.44	6.49	1.75	6.78	1.51
6	7.05	1.59	7.17	1.45	7.24	1.52	7.50	1.28
7	6.81	1.46	6.91	1.36	6.94	1.53	7.21	1.32
8	7.49	1.28	7.61	1.11	7.75	1.28	8.01	0.93
9	6.63	1.81	6.83	1.62	6.18	2.16	6.52	1.90
10	7.13	1.31	7.25	1.16	7.45	1.36	7.71	1.02
11	6.79	1.61	6.97	1.43	7.03	1.78	7.30	1.54
12	6.55	1.79	6.76	1.62	6.71	1.85	6.99	1.61
13	6.97	1.78	7.18	1.47	7.06	1.83	7.28	1.70
14	6.69	1.68	6.91	1.42	7.19	1.34	7.40	1.13
15	6.90	1.54	7.08	1.30	7.19	1.67	7.44	1.42
16	6.79	1.66	6.98	1.46	7.21	1.36	7.41	1.18
17	7.22	1.29	7.34	1.10	7.70	1.13	7.83	0.94
18	6.25	1.77	6.43	1.64	6.95	1.36	7.14	1.17
19	7.98	1.03	8.02	0.96	8.37	0.77	8.49	0.59
20	7.92	1.07	7.98	0.97	8.21	1.06	8.42	0.72
21	7.67	1.14	7.75	1.03	7.65	1.36	7.94	1.03
22	7.58	1.26	7.69	1.07	7.87	1.13	8.07	0.84
23	7.78	1.06	7.81	0.98	8.09	0.96	8.22	0.79
24	7.04	1.40	7.15	1.27	7.33	1.76	7.66	1.45
25	7.15	1.38	7.28	1.26	7.72	1.17	7.92	0.93
26	6.68	1.63	6.83	1.47	7.13	1.46	7.37	1.21
27	7.02	1.25	7.17	1.08	7.63	1.58	7.94	1.25
28	7.58	1.16	7.63	1.12	7.91	1.34	8.12	1.02
29	7.15	1.28	7.23	1.18	7.60	1.41	7.90	1.01
30	6.82	1.44	6.93	1.36	7.15	1.54	7.41	1.30
31	6.95	1.13	7.10	0.96	7.36	1.31	7.67	1.01

APPENDIX VI: Example R software code for performing Generalisability and Decision studies on self-assessment data

- summary(aov(Score~student*Item*patient, data=sa_long))
- formula1 <-
  Score~(1|student)+(1|Item)+(1|patient)+(1|student:Item)+(1|student:patient)+(1|patient:Item)
- g1 <- gstudy(data=sa_long,formula1)
- print(g1)
- d1<-dstudy(g1,colname.objects="student",colname.scores= "Score",data=sa_long)
- print(d1)

APPENDIX VII: Interview protocol (English version) and modifications.

Modifications made to the interview protocol during the interviewing phase: The highlighted questions were added and the ~~strikethrough~~ ones were deleted.

Introduction (5 minutes)

- **Welcome:** Hi [Participant Name], thank you so much for taking the time to chat with me today.
- **Purpose:** This interview is all about understanding your experience with the self-assessment process. We're interested in how you found the experience of evaluating yourself, and learning more about your thinking in that process.
- **Your Rights:** Just so you know, your participation is completely voluntary. You can choose not to answer any question, and you can request to have your data removed any time before July 30.
- **Format:** We'll chat for about 30-45 minutes. No need to be formal – just be yourself!
- **Environment:** It would be great if you could find a relatively quiet spot with minimal distractions for the next hour or so.
- **Review:** Before we dive in, did you have a chance to review your self-assessment forms and patients' forms.
- Do you have any questions or do you have any requests?

General Questions (5 minutes)

- How did you feel about the self-assessment of communication skills experience overall? (probe: could you elaborate?)
- Did the questions in the assessment form made you reflect on aspects you might not have considered otherwise? (Yes, No) (Could you elaborate?)
- ~~How useful did you find the form for assessing yourself meaningfully, and why?~~
- What emotions or thoughts did the self-assessment of communication skills bring up for you? Name feelings
- What impact, if any, do you think the self-assessment process will have on your future interactions with patients?
- What are the negative aspects of your self-assessment experience?

Self-assessment Deep Dive (10-15 minutes)

Thinking Process: Could you walk me through your thought process while you were filling out one of the self-assessment forms (form X3) ? Imagine you're talking out loud as you consider each item. (Probe: What is the difference between 6 and 8 to you? What do they mean to you?)

(Interviewer Note: Listen for whether the participant was trying to guess their score or if they were genuinely reflecting on their performance. Notice different ways of reflections.

Leave room for a breadth of ways they might have been thinking, not just a binary choice (otherwise you might not properly notice variation within those genuinely reflecting)

Alignment between self- and patient assessment (an example of a high performing student with average self-assessment)-(5 minutes)

- After reviewing your self-assessment and the assessments from your patients, what similarities or differences did you notice, and how would you explain them?
- Choose a particular patient: why did your self-assessment differ here the most?
- Choose another patient: Why did you self-assessment was in concordance with the patient assessment here the most?

Possible probing question for the above:

  - *I notice that you scored yourself similarly/differently to your patients on this item, why do you think that was so?*
  - *Were there any times when your patients' assessments surprised you compared to self-assessment? What did you learn from those moments?*

- Your self-assessment in the third encounter was different from fourth encounter? What changed?

Interaction with patients' feedback (5 minutes)

- How would it be different if you did not do self-assessment and just looked at the patients' assessment?
- How did you feel about the patients' feedback, was it useful/meaningful and did you learn anything new from it?

Reflections, Adaptations and interaction with the assessment form (5-minutes)

- According to the consultations recordings, I noticed a difference in your communication style between the first and third patient. Could you tell me more about that?
- Could you tell me of an instance where the assessment instrument guided your actions?

(Interviewer note: possible probing questions: Did the self-assessment form impact your approach to communication? Can you share any instances where it guided your actions?)

- You commented in one form saying " *With difficult patients, I needed to increase my efforts to improve dentist-patient communication*", could you tell me what made you come up with this reflection?
- ~~You answered a number of "why" questions in the comments section, could you please elaborate on how you thought about these questions?~~

- ~~Did you find that your self-awareness of communication with patients changed across interactions with patients 1 to 6? Why not/ why so, do you think?~~
- Did you had a moment of "wait, I should do this" during the consultation session?

Factors affecting self-assessment (5-10 minutes)

- Were you able to gauge how the patient perceived your communication skills? *(probes: If so, how did you do that? What cues did you pick up? Did this change the way you assessed yourself?)*
- Could you mention briefly the factors that affected your self-assessment? Ask about those factors after they answer (stress, acclimating to the assessment, chemistry between you and the patient, previous performance-your ability to determine the diagnosis, your peers)
- Were there any personal expectations or standards you held yourself to that weren't explicitly on the assessment form? (*follow up: Can you say a bit about where this/these came from?*)
- ~~What aspects of self-assessment were challenging/easy? *(probe: what made it easy/difficult?)*~~
- What aspects of self-assessment were challenging to assess? Can you point out certain items?
- What aspects of self-assessment were easy to assess? Can you point out certain items?

Closing (5 minutes)

Before we wrap up, is there anything else you'd like to share about your self-assessment experience that we haven't touched on?

Questions for Me: Do you have any questions for me about this research or anything else?

Confidentiality: I want to reassure you again that everything we've discussed is confidential.

Comments: When I analyse this interview, I'll share any interpretations or conclusions with you so you have a chance to comment.

Thank you again for sharing your valuable insights! This has been very helpful.

APPENDIX VIII: Ethical approval, amendments, participants facing materials, and consent forms

SOCIAL SCIENCES & HUMANITIES
INTERDIVISIONAL RESEARCH ETHICS COMMITTEE
DEPARTMENTAL RESEARCH ETHICS COMMITTEE

Department of Education
15 Norham Gardens, Oxford OX2 6PY
student.curec@education.ox.ac.uk; staff.curec@education.ox.ac.uk



Ghaith Alfakhry ghaith.alfakhry@education.ox.ac.uk
Department of Education, Social Sciences Division
University of Oxford

19 June 2024

Dear Ghaith,

Research ethics approval

Research title: Development and Validation of an Instrument for Self-assessment of Communication Skills in Clinical Dentistry: A Mixed Method Evaluation in Syria

Research ethics reference: EDUC_C1A_24_096

Date of amendment: 19/6/24

Amendment number: 3

The above amendment has been considered on behalf of the Department of Education Departmental Research Ethics Committee (DREC) in accordance with the University's procedures for ethical approval of all research involving human participants.

I am pleased to confirm that, on the basis of the information provided to the DREC, ethics approval has now been granted for this amendment.

Please note the following:

Personal data: It is the responsibility of the PI to ensure that all personal data collected during the project is managed in accordance with the University's guidance and legal requirements.

In-person activities: Any data collection involving in-person interactions with participants must have an up-to-date fieldwork risk assessment in place; further guidance is available from the Safety Office's website.

Amendments: Please notify the committee if you intend to make any further amendments to the information in your ethics application as submitted at date of this approval, as all changes must receive ethical approval prior to implementation. The amendment form is available on the SSH IDREC webpage.

We welcome feedback on your experience of the ethical review process and suggestions for improvement. Please email any comments to staff.curec@education.ox.ac.uk / student.curec@education.ox.ac.uk or ethics@socsci.ox.ac.uk.

Yours sincerely

James Carroll

DREC member
cc: Ariel Lindorff, ariel.lindorff@education.ox.ac.uk

Table 21. A summary of the amendments made to the CUREC 1A form

Amendment 1	
Proposed change	**Reason**
A questionnaire survery will be conducted on 20 patients and 20 student dentists to pretest the clarity and relevance of the items included in the assessment instrument that will be used in the study.	As part of translation, cultural adaptation and linguistic validation of the instrument, it is necessary to pretest the instrument on a sample from the target population to get a sense of its relevance for the purpose it was designed for and the clarity of the phrasing of items
We will not be conducting our study specifically at the Department of Fixed Prosthodontics at Damascus University Dental School.	The study will still take place at Damascus University Dental School; however, we might have to use dental units from other departments depending on availability of these units as my correspondence, Dr. Issam, from the dental school informed me.
Amendment 2	
Proposed change	**Reason**
I would like to request that a research assistant named, Dr. Osama Succari, handle data collection and participants supervision for 7 days from May 7 to May 15. Osama will be responsible for contacting participants, organization and data collection for 7 days. Osama will be trained on how to collect and handle data in accordance to the approved CUREC protocols.	Due to unforeseen circumstances, which were week long national holidays and exams at the dental school, data collection has been delayed. The student researcher is on a Chevening scholarship whose terms do not allow for travelling outside the UK for more than 30 days. Therefore, Ghaith will not be able to stay in Syria to complete data collection and will delegate this responsibility to Dr Osama Succari (research assistant). Assessment forms which Osama is responsible of are completely anonymous (do not contain any identifying information and Osama will be informed of the protocols to keep them anonymous). After transferring the data to an excel sheet, paper forms will be destroyed using a paper shredder.
Amendment 3	
Proposed change	**Reason**
Patients' feedback will be shared with students	For improving communication skills of students with different patients, build confidence, and encourage reflection. Sharing of patients' feedback with students will enable them to cross-analyze them with their own self-assessment which should produce interesting insights in the semi-structured interview.
Visual recording	Students' facial reactions and body language are important to contextualize analysis and to perform naturalistic transcription.

DEPARTMENT OF EDUCATION, CONTACT DETAILS: +44 (0) 1865274024
Ghaith Alfakhry (MSc student)
ghaith.alfakhry@education.ox.ac.uk
Tel: +963 969923534

Dr Ariel Lindorff (supervisor)
Ariel.lindorff@education.ox.ac.uk
Tel: +44 01865 611023

[Self-assessment of Communication Skills in Clinical Dentistry]

STUDENT PARTICIPANTS INFORMATION SHEET

Central University Research Ethics Committee Approval Reference: [EDUC_C1A_24_096]

1. **Introductory paragraph**

You are being invited to take part in a research project. My name is Ghaith Alfakhry, and this is my master's dissertation research at University of Oxford supervised by Dr. Ariel Lindorff. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part.

2. **Why is this research being conducted?**

Communication skills are universally recognized as a fundamental clinical skill. Our research aims to develop a tool for dental trainees to self-assess their communication skills to find areas for improvement.

3. **Why have I been invited to take part?**

This research targets dental students in their 5th year stage at the Faculty of Dental Medicine, Damascus University. The age range for student participants is between 20 and 25. Participants are selected randomly from the pool of students who initially agree to participate in the screening survey.

4. **Do I have to take part?**

No. It is up to you to decide whether to take part. You can withdraw yourself from the research, without giving a reason, by advising me of this decision. The deadline by which you can withdraw any information you have contributed to the research is May 31, 2024; at this point, all data will have been pseudonymised and included in the analysis.

5. **What will happen to me if I take part in the research?**

The research will take place at Faculty of Dental Medicine, Damascus University. Informed consent will be taken beforehand. You will be involved in the research from mid-April to May 31, 2024. You will be asked to attend to six patients as if they were patients visiting your clinic; you will only clinically examine them and will not conduct any invasive procedure. After each patient, you will be instructed to complete a self-assessment form of communication skills. Patient participants will also have to complete a form to assess the communication performance of their student dentist. There are three standardized patients and three real patients; you will not be informed which is which so that does not affect your behaviour when you interact with standardized patients, your interaction will be recorded; however, you have the right refuse having your interactions recorded if you wish; your interactions with real patients will not be recorded. You will be asked to complete a questionnaire after you examine all 6 patients. You may be asked to participate in an video recorded

Online interview (Via Microsoft Teams) which you can say no to even if you decide to participate now.

6. **What are the possible disadvantages and risks in taking part?**

It is important for you to consider the time commitment and effort required to examine six patients, which should take three hours in total) . The assessment conducted as part of the research will remain de-identified and will not affect your grades.

7. **Are there any benefits in taking part?**

Research participants can expect to receive the supervisor and patients' feedback on their communication skills at the end of the day. A certificate of participation will be provided. A copy of the self-assessment form will be shared with students for future use which can be beneficial for exam preparation and clinical practice.

8. **What information will be collected and why is the collection of this information relevant for achieving the research objectives?**

Name and contact details will be collected and afterwards your name will be replaced with a numeric code which you will use during the study. Your self-assessment on the given form will be collected as well as your answers on the survey and the interview answers and recording. These will help us understand how dental students self-assess themselves using our assessment form. You will not be identified and no personal data will be shared in any research outputs from the project.

9. **Will the research be published? Could I be identified from any publications or other research outputs?**

The findings from the research will be written up in a dissertation and academic publication and conference presentations; data will be shared to support publication in an academic journal. You will not be directly identifiable from these outputs; you will be pseudonymized (your name will be replaced with a unique code). A copy of my thesis/ dissertation will be deposited both in print and online in the Oxford University Research Archive where it will be publicly available to facilitate its use in future research.

10. **Data Protection**

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from the University's Information Compliance web site at https://compliance.admin.ox.ac.uk/individual-rights.

11. **Who has reviewed this research?**

This research has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: **xxxxx**).

12. **Who do I contact if I have a concern about the research or I wish to complain?**

If you have a concern about any aspect of this research, please contact [Ghaith Alfakhry, Email: ghaith.alfakhry@education.ox.ac.uk.) and I will do our best to answer your query. You may also contact my supervisor, Dr Ariel Lindorff (ariel.lindorff@education.ox.ac.uk) if you prefer. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt

with. If you wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:
The Chair: Dr Elizabeth Wonnacott. Email: student.curec@education.ox.ac.uk; Address: University of Oxford, 15 Norham Gardens, Oxford OX2 6PY.
The Chair, Social Sciences & Humanities Interdivisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Boundary Brook House, Churchill Drive, Headington, Oxford OX3 7GB

13. Further Information and Contact Details

If you would like to discuss the research with someone beforehand, please contact: Ghaith Alfakhry, Department of Education, University of Oxford. Phone number in Syria: 0969923534.

DEPARTMENT OF EDUCATION, CONTACT DETAILS: +44 (0) 1865274024
Ghaith Alfakhry (MSc student)
ghaith.alfakhry@education.ox.ac.uk
Tel: +963 969923534

Dr Ariel Lindorff (supervisor)
Ariel.lindorff@education.ox.ac.uk
Tel: +44 01865 611023

Student Participants consent to take part in [Self-assessment of Communication Skills in Clinical Dentistry]

Central University Research Ethics Committee (CUREC) approval reference: EDUC_C1A_24_096
Purpose of Study: This research aims to pilot test an assessment tool for dental trainees to self-assess communication skills with patients in clinical settings.

	Please initial each box if you agree with the statement
I confirm that I have read and understand the information sheet for the above research. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
I understand that my participation is voluntary and that I am free to withdraw at any point [until **31/May/2024**], without giving any reason.	☐
I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	☐
I understand I will not be identifiable from any publications or conference presentations.	☐
I consent to being audio-recorded in my interactions with standardized patients (actors)	☐
I consent to being audio recorded if I was selected to be interviewed.	☐
I consent to being video recorded if I was selected to be interviewed.	☐
I understand how audio recordings will be used in research outputs.	☐
Use of quotations: Please indicate your preference (select *one* option): a) I do not wish to be quoted. **or** b) I agree to the use of quotations in research outputs if I am not identifiable.	☐ ☐
I give permission for you to contact me again to clarify information.	☐

I understand how to raise a concern or make a complaint.	☐
I agree to take part.[1]	☐
I agree that my personal contact details can be retained in a secure database so that the researchers can contact me to clarify information should that be necessary	YES / NO

_____	dd / mm / yyyy	_____
Name of participant	Date	Signature

_____	dd / mm / yyyy	_____
Name of person taking consent	Date[2]	Signature

DEPARTMENT OF EDUCATION, CONTACT DETAILS: +44 (0) 1865274024
Ghaith Alfakhry (MSc student)
ghaith.alfakhry@education.ox.ac.uk
Tel: +963 969923534

Dr Ariel Lindorff (supervisor)
Ariel.lindorff@education.ox.ac.uk
Tel: +44 01865 611023

[Self-assessment of Communication Skills in Clinical Dentistry]

PATIENT PARTICIPANTS INFORMATION SHEET

Central University Research Ethics Committee Approval Reference: [EDUC_C1A_24_096]

1. **Introductory paragraph**

You are being invited to take part in a research project. My name is Ghaith Alfakhry, and this is my master's dissertation research at University of Oxford supervised by Dr. Ariel Lindorff. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part.

2. **Why is this research being conducted?**

Communication skills are universally recognized as a fundamental clinical skill. Our research aims to develop a tool for dental trainees to self-assess their communication skills to find areas for improvement.

3. **Why have I been invited to take part?**

The research aims to have dental trainees assess their communication performance with 6 patients for educational purposes and patients are necessary for this research project.

4. **Do I have to take part?**

No. It is up to you to decide whether to take part. You can withdraw yourself from the research, without giving a reason, by advising me of this decision. The deadline by which you can withdraw any information you have contributed to the research is May 31, 2024; at this point, all data will have been pseudonymised and included in the analysis.

5. **What will happen to me if I take part in the research?**

The research will take place at the Faculty of Dental Medicine, Damascus University. Informed consent will be taken beforehand. You will be requested to be at the clinics to be clinically examined by 60 students over ten days (6 students per day), starting from mid-April. You can ask to pause or stop the research activities at any time. If you were chosen to be in the standardized patients group, you will receive assessment training of four hours over two days before you are examined by student dentists; further, your interactions with the student dentists will be recorded if you were chosen as standardized patient assuming that both you and the student participant consented to this. You will need to fill an assessment form for each student who examines you; filling the assessment form should take 10 minutes maximum.

6. **What are the possible disadvantages and risks in taking part?**

You will go through 60 clinical oral examinations which can be tiring. That is why these will be divided up to 10 days. You will be given intermittent breaks during the one day to avoid jaw fatigue. As for infection control, supervisors will make sure dental trainees adhere to the highest standard of practice.

7. **Are there any benefits in taking part?**

You will be given a certificate for participation and will be compensated with 4000 Syrian pounds for each clinical examination you undertake amounting to a total of 240,000 Syrian for all 60 planned clinical examinations at the end of the 10th day. Standardized patients, who are dentists, can also expect to benefit from training in assessing communication skills. The assessment training will be reflected in your certificate of participation.

8. **What information will be collected and why is the collection of this information relevant for achieving the research objectives?**

Name and contact details will be collected and afterwards your name will be replaced with a numeric code which you will use during the study. Your feedback in the assessment forms regarding student participants communication performance will be collected to help us analyse how accurate students are in self-assessing their communication approach which is the research objective. You will not be identified and no personal data will be shared in any research outputs from the project.

9. **Will the research be published? Could I be identified from any publications or other research outputs?**

The findings will be written up in a dissertation, academic publication, and conference presentations; data will be shared to support publication in an academic journal. You will not be directly identifiable from the outputs; you will be pseudonymized (your name will be replaced with a unique code). A copy of my dissertation will be deposited both in print and online in the Oxford University Research Archive where it will be publicly available to facilitate its use in future research.

10. **Data Protection**

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from the University's Information Compliance web site at:
https://compliance.admin.ox.ac.uk/individual-rights.

11. **Who has reviewed this research?**

This research has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: **xxxxx)**.

12. **Who do I contact if I have a concern about the research or I wish to complain?**

If you have a concern about any aspect of this research, please contact [Ghaith Alfakhry, Email: ghaith.alfakhry@education.ox.ac.uk.) and I will do our best to answer your query. You may also contact my supervisor, Dr Ariel Lindorff (ariel.lindorff@education.ox.ac.uk) if you prefer. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:
The Chair: Dr Elizabeth Wonnacott. Email: student.curec@education.ox.ac.uk; Address: University of Oxford, 15 Norham Gardens, Oxford OX2 6PY.
The Chair, Social Sciences & Humanities Interdivisional Research Ethics Committee;
Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Boundary Brook House, Churchill Drive, Headington, Oxford OX3 7GB

13. Further Information and Contact Details

If you would like to discuss the research with someone beforehand, please contact: Ghaith Alfakhry, Department of Education, University of Oxford. Phone number in Syria: 0969923534.

DEPARTMENT OF EDUCATION, CONTACT DETAILS: +44 (0) 1865274024
Ghaith Alfakhry (MSc student)
ghaith.alfakhry@education.ox.ac.uk
Tel: +963 969923534

Dr Ariel Lindorff (supervisor)
Ariel.lindorff@education.ox.ac.uk
Tel: +44 01865 611023

Patient participants consent to take part in [Self-assessment of Communication Skills in Clinical Dentistry]

Central University Research Ethics Committee (CUREC) approval reference: **EDUC_C1A_24_096**
Purpose of Study: This research aims to test an assessment tool for dental trainees to self-assess communication skills with patients in clinical settings.

	Please initial each box if you agree with the statement
I confirm that I have read and understand the information sheet for the above research. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
I understand that my participation is voluntary and that I am free to withdraw at any point [until **31/May/2024**], without giving any reason.	☐
I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	☐
I understand I will not be directly identifiable from any publications or conference presentations.	☐
I consent to being audio-recorded in my interactions students should I be a standardized patient (if you were chosen as a real patient, please ignore this item)	☐
Use of quotations: Please indicate your preference (select *one* option): a) I do not wish to be quoted. **or** b) I agree to the use of quotations in research outputs if I am not directly identifiable.	☐ ☐
I give permission for you to contact me again to clarify information.	☐
I understand how to raise a concern or make a complaint.	☐
I agree to take part.[1]	☐

I agree that my personal contact details can be retained in a secure database so that the researchers can contact me to clarify information should that be necessary	YES / NO

_____	dd / mm / yyyy	_____
Name of participant	Date	Signature
_____	dd / mm / yyyy	_____
Name of person taking consent	Date[2]	Signature

[2] To be signed and dated in the presence of the participant. Once this has been signed by both parties the participant should receive a copy of the signed and dated participant consent form. The original signed and dated consent form should be kept with the project's main documents, which must be kept in a secure location.

DEPARTMENT OF EDUCATION, CONTACT DETAILS: +44 (0) 1865274024
Ghaith Alfakhry (MSc student)
ghaith.alfakhry@education.ox.ac.uk
Tel: +963 969923534

Dr Ariel Lindorff (supervisor)
Ariel.lindorff@education.ox.ac.uk
Tel: +44 01865 611023

[Self-assessment of Communication Skills in Clinical Dentistry]

PILOT TESTING PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee Approval Reference: [EDUC_C1A_24_096]

1. **Introductory paragraph**

You are being invited to take part in a research project. My name is Ghaith Alfakhry, and this is my master's dissertation research at University of Oxford supervised by Dr. Ariel Lindorff. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part.

2. **Why is this research being conducted?**

Communication skills are universally recognized as a fundamental clinical skill. Our research aims to develop a tool for dental trainees to self-assess their communication skills to find areas for improvement.

3. **Why have I been invited to take part?**

The research aims to develop and validate an instrument to assess dental students communication skills. You are being invited to take part to test the clarity and relevance of the items used in the assessment instrument.

4. **Do I have to take part?**

No. It is up to you to decide whether to take part. As the questionnaire survey will be anonymous, it will unfeasible for you to withdraw the data you offered.

5. **What will happen to me if I take part in the research?**

You will need to complete a questionnaire survey on the clarity and relevance of an assessment instrument designed to assess dental students communication skills in clinical settings. Filling the form should take approximately 10 minutes.

6. **What are the possible disadvantages and risks in taking part?**

As the data you will offer will be anonymous, the risk of participating in this research is effectively zero. We are mindful of your time and truly appreciate your help in this research.

7. **Are there any benefits in taking part?**

There are no direct benefits to you; however, you will be participating in a research that aims to improve communication skills of student dentists at this dental school which can reflect positively on future patients and dentists.

8. **What information will be collected and why is the collection of this information relevant for achieving the research objectives?**

Only relevant personal non-identifying information will be collected such as sex and age will be collected.

9. **Will the research be published? Could I be identified from any publications or other research outputs?**

The findings will be written up in a dissertation, academic publication, and conference presentations; data will be shared to support publication in an academic journal. You will not be identifiable from the outputs. A copy of my dissertation will be deposited both in print and online in the Oxford University Research Archive where it will be publicly available to facilitate its use in future research.

10. **Data Protection**

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from the University's Information Compliance web site at:
https://compliance.admin.ox.ac.uk/individual-rights.

11. **Who has reviewed this research?**

This research has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: **xxxxx**).

12. **Who do I contact if I have a concern about the research or I wish to complain?**

If you have a concern about any aspect of this research, please contact [Ghaith Alfakhry, Email: ghaith.alfakhry@education.ox.ac.uk.) and I will do our best to answer your query. You may also contact my supervisor, Dr Ariel Lindorff (ariel.lindorff@education.ox.ac.uk) if you prefer. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:
The Chair: Dr Elizabeth Wonnacott. Email: student.curec@education.ox.ac.uk; Address: University of Oxford, 15 Norham Gardens, Oxford OX2 6PY.
The Chair, Social Sciences & Humanities Interdivisional Research Ethics Committee;
Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Boundary Brook House, Churchill Drive, Headington, Oxford OX3 7GB

13. **Further Information and Contact Details**

If you would like to discuss the research with someone beforehand, please contact: Ghaith Alfakhry, Department of Education, University of Oxford. Phone number in Syria: 0969923534.

DEPARTMENT OF EDUCATION, CONTACT DETAILS: +44 (0) 1865274024
Ghaith Alfakhry (MSc student)
ghaith.alfakhry@education.ox.ac.uk
Tel: +963 969923534

Dr Ariel Lindorff (supervisor)
Ariel.lindorff@education.ox.ac.uk
Tel: +44 01865 611023

Pilot testing participants consent to take part in [Self-assessment of Communication Skills in Clinical Dentistry]

Central University Research Ethics Committee (CUREC) approval reference: **EDUC_C1A_24_096**
Purpose of Study: This research aims to test an assessment tool for dental trainees to self-assess communication skills with patients in clinical settings.

	Please initial each box if you agree with the statement
I confirm that I have read and understand the information sheet for the above research. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
I understand that my participation is voluntary and that I will not be able to withdraw the data I have offered as it will anonymous.	☐
I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	☐
I understand I will not be identifiable from any publications or conference presentations.	☐
Use of quotations: Please indicate your preference (select *one* option): a) I do not wish to be quoted. **or** b) I agree to the use of quotations in research outputs if I am not directly identifiable.	☐ ☐
I give permission for you to contact me again to clarify information.	☐
I understand how to raise a concern or make a complaint.	☐
I agree to take part.[1]	☐

_____	dd / mm / yyyy	_____
Name of participant	Date	Signature
_____	dd / mm / yyyy	_____
Name of person taking consent	Date[2]	Signature

[2] To be signed and dated in the presence of the participant. Once this has been signed by both parties the participant should receive a copy of the signed and dated participant consent form. The original signed and dated consent form should be kept with the project's main documents, which must be kept in a secure location.

APPENDIX IX: Rotated factor loadings for the first two factors with direct oblimin rotation

Pattern Matrix[a]

	Factor	
	1	2
s18	.839	
s14	.783	
s16	.754	
s13	.716	
s12	.716	
s15	.693	
s3	.627	
s5	.620	
s26	.620	
s30	.567	
s4	.566	
s27	.561	
s7	.537	
s10	.491	
s25	.469	-.409
s11	.462	
s29	.428	
s9	.423	
s23		-.903
s20		-.887
s19		-.886
s21		-.773
s2		-.659
s8		-.647
s22		-.638
s24	.426	-.448
self-assessment: item 1		-.430
s28		-.425
s17		-.416
s6		-.403

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 22 iterations.

APPENDIX X: Descriptive statistics, skewness and kurtosis, and tests of normality for aggregated 30 items of self-assessment and patient assessment (ssum30=aggregated SA, psum30=aggregated PA).

Descriptives

			Statistic	Std. Error
ssum30	Mean		212.5463	1.72459
	95% Confidence Interval for Mean	Lower Bound	209.1535	
		Upper Bound	215.9391	
	5% Trimmed Mean		213.4966	
	Median		213.0000	
	Variance		963.642	
	Std. Deviation		31.04258	
	Minimum		95.00	
	Maximum		270.00	
	Range		175.00	
	Interquartile Range		45.00	
	Skewness		-.455	.135
	Kurtosis		.418	.270
psum30	Mean		220.5556	1.67931
	95% Confidence Interval for Mean	Lower Bound	217.2518	
		Upper Bound	223.8593	
	5% Trimmed Mean		222.0919	
	Median		222.5000	
	Variance		913.709	
	Std. Deviation		30.22762	
	Minimum		125.00	
	Maximum		270.00	
	Range		145.00	
	Interquartile Range		43.00	
	Skewness		-.620	.135
	Kurtosis		.267	.270

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ssum30	.048	324	.064	.977	324	<.001
psum30	.056	324	.015	.966	324	<.001

a. Lilliefors Significance Correction

Appendix XI: Mixed repeated measures ANOVA for SA and PA for the defined three groups based on patient assessment tiers.

Generally, SA scores were lower than PA scores across all tiers. Between-subject effects revealed a significant difference in SA between groups, with post hoc analysis showing a significant difference between the low and high-tier groups ($P = 0.015$). The between-subject effect for PA was highly significant, with a very large effect size ($P < 0.001$, Partial $\eta^2 = 0.999$), which is expected since these groups were defined by ranking the aggregated PA scores across all six patients for each student (**see section 3.5.4**).

Table 22. Mixed repeated measures ANOVA for self-assessment and patient assessment for the defined three groups based on Patient assessment tiers.

PA tier		Self-assessment			Patient assessment		
Low-tier	N	Mean	SD		Mean	SD	
1st patient	17	184.1	25.1		199.1	30.6	
2nd patient	17	203.6	26.3		213.8	27.3	
3rd patient	17	201.7	40.4		210.2	30.4	
4th patient	17	198.8	40.9		198.5	32.7	
5th patient	17	203.4	31.8		197.4	34.9	
6th patient	17	202.9	30.4		214.2	36.4	
Mid-tier							
1st patient	19	196.9	34.0		195.3	25.7	
2nd patient	19	207.6	32.1		225.2	22.3	
3rd patient	19	218.6	28.0		224.2	24.7	
4th patient	19	218.4	29.4		224.7	24.3	
5th patient	19	217.6	27.5		233.0	25.4	
6th patient	19	218.7	29.5		225.4	25.5	
High-tier							
1st patient	18	215.1	26.7		215.3	31.8	
2nd patient	18	224.2	23.5		235.8	18.7	
3rd patient	18	223.5	27.5		233.6	18.4	
4th patient	18	231.7	23.1		240.9	24.2	
5th patient	18	228.6	21.9		236.1	21.5	
6th patient	18	225.9	19.3		242.1	24.7	
Between subject effects:		P-value	Partial η^2	Post hoc	*P*-value	Partial η^2	Post hoc
		0.019	0.144	ht>lt*	<0.001	0.999	ht>lt*** ht>mt*** mt>lt***

Ht=high-tier; mt=mid-tier; lt=low-tier.