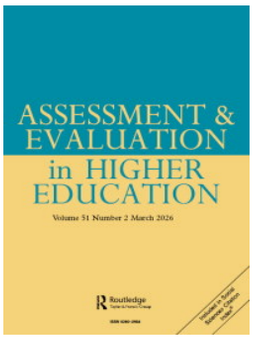

Routledge
Taylor & Francis Group

Assessment & Evaluation in Higher Education

ISSN: 0260-2938 (Print) 1469-297X (Online) Journal homepage: www.tandfonline.com/journals/caeh20

Interview assessment as compressed judgement: how evidence becomes recognisable under interactional constraint

Emma J. Folwell

To cite this article: Emma J. Folwell (10 Apr 2026): Interview assessment as compressed judgement: how evidence becomes recognisable under interactional constraint, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2026.2656293

To link to this article: https://doi.org/10.1080/02602938.2026.2656293

Published online: 10 Apr 2026.

ASSESSMENT & EVALUATION IN HIGHER EDUCATION
https://doi.org/10.1080/02602938.2026.2656293

Routledge
Taylor & Francis Group

Interview assessment as compressed judgement: how evidence becomes recognisable under interactional constraint

Emma J. Folwell

Arts, Humanities and Human Sciences, Birmingham Newman University, Birmingham, UK

ABSTRACT

Assessment judgement research recognises evaluative decisions as interpretive and situated, yet has paid less attention to how evidential sufficiency becomes recognisable when judgement is enacted in real time. This paper examines interview-based assessment as an assessment instrument whose significance lies in the judgement logic it operationalises. Drawing on reflexive thematic analysis of qualitative interviews with staff and students on an undergraduate programme employing assessed interviews, the study examines how evidence is organised, stabilised, and recognised under conditions of temporal and interactional constraint. It shows how interview questions function as pre-authorising devices that delimit what can count as legitimate evidence in advance, how evidential adequacy is produced interactionally rather than accumulated, and how student preparation becomes oriented towards recognisability under compression. These dynamics are explored as compressed judgement: a form of assessment judgement oriented towards decision-readiness within a bounded and irreversible encounter.

KEYWORDS

Assessment judgement; interview-based assessment; oral assessment

Introduction

Interviews and oral examinations are commonly discussed as pedagogical encounters: sites of dialogue, authenticity, anxiety, or performance. Within this framing, interviews are typically treated as methods for eliciting evidence or evaluating individual competence (Nelson 2010), with attention directed towards student experience, affect, or preparation. Recent interest in these formats has intensified as institutions seek assessment options perceived as less vulnerable to generative artificial intelligence than traditional written assignments (Moorhouse, Yeo, and Wan 2023; Sullivan, Kelly, and McLaughlan 2023). What remains comparatively under-examined is how interview-based assessment operates as a judgement system: how evidential sufficiency becomes recognisable, and how summative decisions are enacted in real time under

CONTACT Emma J. Folwell e.folwell@newman.ac.uk Arts, Humanities and Human Sciences, Birmingham Newman University, Birmingham, UK.

interactional constraint. Without such an account, attention shifts towards student performance rather than the assessment architecture that organises what counts as evidence.

Existing research often explores preparation, feedback, self-efficacy, and affective factors, alongside design interventions intended to improve reliability or validity. Interview assessment is frequently justified as offering deeper or more authentic access to student understanding, with anxiety commonly treated as a mediating variable to be managed (Stephenson, Johnson-Glauch, and Cruchley 2025). While this work has generated valuable insights into student experience and assessment design, it tends to leave the operation of judgement itself comparatively opaque, particularly how evidential adequacy becomes recognisable in the moment of assessment.

This paper approaches interview-based assessment differently. Rather than treating the interview as a pedagogical interaction or performance event, it conceptualises it as an assessment instrument whose significance lies in the judgement logic it enacts. The focus is not on student experience, but on how assessment formats organise recognition: how articulation, responsiveness, and self-correction become legible as evidence under conditions of immediacy and interaction. In interview assessment, judgement does not follow the accumulation of evidence but is produced through real-time conversational uptake within a bounded and irreversible encounter.

Drawing on reflexive thematic analysis of qualitative research interviews with staff and students in an undergraduate programme employing assessed interviews, the paper advances a judgement-theoretical account of interview assessment. It examines how interview questions operate as design devices that delimit what can count as a legitimate answer in advance, how evidential adequacy is produced and stabilised through interaction, and how student preparation becomes oriented towards recognisability rather than authenticity. The paper offers a foundation for understanding how interview formats structure recognition at the point of judgement, extending judgement-theoretical accounts by examining how evidential sufficiency is enacted under conditions of interactional and temporal compression.

Literature review

Oral examinations, vivas, and structured interview assessments have a long history, particularly in professional training (Joughin 1998). Oral formats are frequently justified as offering more diagnostic or conceptually revealing assessments of student learning, and are often framed as more authentic where reasoning or professional judgement are at stake (Harsy 2024; Nelson 2010). Oral examinations and structured interview assessments differ from presentation-based assessment because they are assessor-led and question-driven, not student-delivered performances. In assessed interviews, students must respond and adapt in real time to prompts, follow-up questions, and shifts in direction initiated by assessors, typically under tight time constraints. Research on cumulative oral exams (Crecelius, DeRuisseau, and Brandauer 2021), oral case presentations (Sox et al. 2018), and online oral assessments (McBain et al. 2016) shows that these formats are interactive and unscripted, even though they vary by discipline and surface genre. These interactional features not only

distinguish oral assessments pedagogically, but also shape the conditions under which evidence is produced and recognised at the point of judgement.

Empirical studies have explored the value of design interventions, including structured practice opportunities such as mock oral examinations, in supporting student preparation for oral assessment under examination conditions (Fischer et al. 2016; Van Ginkel et al. 2015). Feedback, particularly from assessors, has been repeatedly identified as a critical factor, with studies indicating differential effects depending on feedback source, timing, and structure (Dickson, Harvey, and Blackwood 2019; Van Ginkel et al. 2017). Within this framing, feedback is typically theorised as a mechanism for improving future performance by clarifying expectations, identifying weaknesses, and supporting skill development (Hattie and Timperley 2007). Self-reflection and metacognitive strategies also feature prominently alongside feedback as means of supporting preparation for oral performance (Oliveira et al. 2021). Across the literature, oral assessment is often treated as a designable practice whose limitations can be mitigated through targeted interventions focused on preparation, feedback, and self-regulation (Stephenson, Johnson-Glauch, and Cruchley 2025). Judgement is largely treated as responsive to improved performance, rather than as a process with its own operative logic.

Research has also identified anxiety and affective load as salient features of oral examination contexts. Comparative studies indicate that oral examinations are associated with higher test anxiety than written assessments, with anxiety in oral contexts more strongly related to social exposure and evaluative visibility (Laurin-Barantke et al. 2016). Anxiety is often treated as a factor that may mediate performance quality, prompting a focus on interventions intended to support students' affective orientation to oral assessment. In this literature, anxiety is typically framed as an obstacle to be mitigated, rather than as an indicator of the exposure and irreversibility of judgement in oral assessment.

Oral assessments have been suggested as offering deeper or more diagnostic access to student understanding, particularly when compared with written responses, although such claims typically rest on assumptions about how spoken interaction functions as evidence at the point of judgement. Studies comparing written examination performance with responses elicited through oral questioning suggest that written scores do not always align with conceptual understanding, with oral interviews exposing misconceptions that remain obscured in written responses (Sato, Hill, and Lo 2019). Comparative studies using equivalent questions have also reported higher mean performance in oral assessments than in written formats, alongside student perceptions that oral assessment is more useful despite higher reported anxiety (Huxham, Campbell, and Westwood 2012). Oral exams are deployed in some domains specifically to probe conceptual understanding where procedural correctness alone is insufficient (Harsy 2024). These works support claims about the pedagogical value of oral assessment, while also highlighting how spoken interaction actively reconfigures what can count as evidence.

There is a tendency to locate the challenges and successes of oral assessment at the level of student performance or preparation, with judgement itself often treated as either a measurement problem addressed through rubrics and reliability procedures (Ragupathi and Lee 2020) or as a consequence of improved performance.

Recent work frames oral assessment primarily in terms of validity, integrity and student performance, with design features proposed as mitigations for recognised limitations (e.g. Nallaya et al. 2024). What remains comparatively under-explored is how evidence becomes recognisable and actionable for summative judgement.

Assessment research conceptualises judgement as the interpretive act through which assessors determine whether available evidence is sufficient relative to standards. This determination exceeds mechanical application of explicit criteria, particularly in complex tasks where quality cannot be exhaustively specified in advance (Sadler 2009). While evaluative judgement is often used to describe students' capacity to judge the quality of work, judgement in assessment practice more commonly refers to the professional decision-making through which evidential adequacy is recognised. Scholarship has shown this activity to be socially learned and institutionally patterned, where standards become meaningful through shared recognition rather than formal specification alone (Rust, Price, and O'donovan 2003; Shay 2005). Empirical studies further suggest that criteria and documentation often stabilise and justify decisions retrospectively, rather than govern recognition continuously during assessment encounters, particularly under practical constraint (Bloxham, Boyd, and Orr 2011; Ecclestone 2001).

Here the term judgement logic is used to refer to the patterned conditions under which evidential adequacy becomes recognisable and decision-ready within a given assessment format. Judgement logic concerns how assessment architectures organise recognition: what can count as sufficient, how sufficiency is stabilised, and how decisions become actionable under constraint. In assessor-led interview assessment, these conditions are especially acute. The encounter is time-bounded, interactive and irreversible, requiring assessors to reach decisions without exhaustive evidential exploration or later correction. While existing scholarship explains why judgement cannot be fully governed by criteria, it has attended less directly to how particular contributions come to be recognised as evidentially sufficient within live assessment interactions.

Methodology

This paper adopts a qualitative, interpretive design informed by reflexive thematic analysis (RTA) as articulated by Braun and Clarke (2022). RTA was used to interrogate how assessed interviews operate as judgement formats. Although coding was initially inductive, analytic development proceeded by reworking codes through a judgement-theoretical lens. Rather than aggregating views or practices, codes were interrogated for what they revealed about the conditions under which evidential adequacy became recognisable in real time. Themes were therefore formulated as claims about judgement logic rather than as summaries of participant experience. Analysis focused on latent assumptions about legitimacy and recognition that structured how interview assessment was understood. Because the analysis targets judgement conditions and recognition logics rather than interactional mechanics, participant accounts are treated as analytically appropriate sites for examining how evidential adequacy is understood and anticipated across interview assessments, rather than as substitutes for observational data.

This study draws on interview data generated as part of a broader qualitative project examining staff and student experiences of learning, teaching and assessment

within an undergraduate programme. The dataset comprises semi-structured interviews with 13 students and 4 staff members, each lasting between 30 min and one hour. Student interviews were conducted at the beginning of Level Six study, after participants had completed assessed interviews at Levels Four and Five. Staff interviews explored experiences of assessment design, marking and moderation across the programme. Informed consent was gained for participation in the interviews, which were conducted and recorded *via* Microsoft Teams. Interviews were transcribed verbatim and anonymised, with all identifying information removed. To protect participants within a small and identifiable institutional context, the programme site is also anonymised. Staff interview transcripts were labelled ST1-4, and student transcripts SP1-13. Ethical approval for the original study included consent for further research use of the data. Secondary analysis was undertaken as a theoretically driven re-engagement with the dataset, enabling exploration of latent dimensions of judgement not exhausted by the initial analytic focus.

Both staff and student accounts were included because the study examines judgement as a relational process rather than a unilateral act of marking. While staff enact summative decisions, students orient themselves towards anticipated judgement conditions in preparation and response. Student accounts therefore illuminate how recognition is anticipated and navigated, while staff accounts show how evidential sufficiency is interpreted and stabilised. The small number of staff participants reflects the size of the assessment team within the programme context.

Programme context

The assessed interviews examined here form part of a longitudinal, accelerated interdisciplinary undergraduate programme that employs criterion-referenced ipsative assessment. At each level, students complete an assessed portfolio early in the level and an assessed interview later in the level. The Level Four interview is a 20-minute discussion focused on lifewide learning and personal development, with questions provided in advance. The Level Five interview is a 25–30 min graduate job interview simulation exploring academic and professional development. Questions are not provided in advance. All interviews are conducted by two assessors, double marked and group moderated. The programme is structured around a taxonomy of five non-hierarchical categories, which provides a stable frame from which interview assessment marking criteria are derived. Assessors recorded grades independently before moderation, and marking criteria were used to structure post-interview deliberation. However, criteria were not operationalised as question-level checklists during the assessed interview. While standards were formally specified, their enactment depended on real-time interpretive recognition rather than explicit criterion matching. Alongside this summative assessment, students participate in monthly one-to-one academic life coaching sessions that support dialogic feedback, student-led goal setting, and reflective self-audit across the programme taxonomy. While the programme incorporates ipsative elements, the analysis in this paper does not examine learning progression over time. Instead, it focuses on how judgement is enacted within individual interview encounters, treating each interview as a bounded site of evidential recognition.

Researcher positioning

The researcher occupies a dual insider role as head of subject and co-creator of the programme, with line management responsibility for staff participants and prior pedagogical and assessment relationships with some student participants. The researcher also conducted all interviews. This proximity creates both analytic opportunity and ethical responsibility. Data collection was deliberately timed to mitigate coercive pressure: interviews were conducted after Level Five decisions and several months before the next submission point.

Insider knowledge was treated as a situated analytic resource, shaping sensitivity to the nuances of assessment practice and participant language. RTA provided a framework for interrogating rather than normalising familiar practices. Particular attention was paid to moments where data unsettled the researcher's pre-existing assumptions. The researcher initially assumed that interview assessments would be experienced primarily as anxiety-inducing, and that staff shared relatively clear understandings of what was being assessed. These assumptions were disrupted by the data, which revealed more ambivalent student experiences and persistent ambiguity in staff accounts of judgement. Rather than resolving these tensions, the analysis foregrounded them.

Analysis followed the recursive and reflexive phases of RTA, beginning with familiarisation and inductive coding across the dataset. Codes were then grouped into five meaning clusters. Codes relating to nervousness, fluency, preparation, and composure initially sat alongside one another, inviting interpretation as individual confidence or communication skill. Rather than resolving these patterns into trait-based themes, the analysis repeatedly reworked them across clusters as judgement-relevant cues that become salient under time-pressured and ambiguous assessment conditions. The analysis deliberately avoids claims about learning gain, skill acquisition, or employability outcomes, which fall beyond the analytic scope of the dataset. Similarly, repeated references to 'having examples ready' were not treated as evidence of experiential learning per se, but as interactionally legible artefacts through which credibility could be recognised in an interview context.

As the study focuses on assessment judgement, an inherently evaluative and consequential practice, insider reflexivity was treated as an ethical as well as analytic responsibility. Given the researcher's insider position and the focus on judgement, particular care was taken to avoid deficit framings of student accounts. Expressions of anxiety or extensive preparation were not coded as individual shortcomings, but interpreted as rational responses to the structural conditions of assessed interviews. No demographic data were collected due to the risk of deductive disclosure within a small programme. As a result, the analysis does not make claims about differential impacts.

Findings

Conditions of judgement in assessed interviews

Judgement emerged in the dataset as configured by three structural conditions: temporality, ambiguity, and relationality. These conditions shape what is recognised

as legitimate performance. Temporality denotes a bounded decision window in which assessors evaluate responses in real time. Distinct from interview duration, this decision window is irreversible, forcing selective translation of expansive experience into assessable evidence. Because only part of a student's knowledge is visible within the interaction, students prioritise and structure responses to maximise immediate intelligibility: 'I mean… it's 30 min… if you're really enthusiastic about something, you can talk about it forever, can't you?' (SP6). What becomes assessable is what can be communicated coherently and recognised quickly.

Ambiguity arises because criteria are not fully operationalised during the interview itself, leaving expectations underspecified in practice. Criteria recede in the interaction but continue to structure recognition indirectly, shifting standards towards tacit assessor interpretation rather than explicit criterion application. As one student noted, 'The marking criteria, I understand the words, but I don't always know how to apply them to my life yet, or my uni work' (SP11). The issue was not lack of access to standards, but difficulty translating formally specified criteria into interactionally usable evidence. Students navigate uncertainty about recognition through iterative exposure and feedback across encounters rather than through written criteria alone. As one student observed, 'Sometimes you will just never know how you're going to answer a question until you get feedback' (SP13), indicating that ambiguity redistributes rather than suspends standards.

Relationality captures the interactional emergence of judgement through assessor uptake and its stabilisation through collective moderation. Coaching relationships further shape how experience becomes articulable and recognisable, extending relational contingency beyond the interview interaction. Student accounts show that answers only acquire evidential status when assessors acknowledge and integrate them into the conversational flow. Staff similarly locate credibility in communal calibration: 'that's why it's so good that there's four or five of us [marking and moderating]… that's where you get the credibility' (ST4). Relational infrastructures enable the recognition of evidential adequacy while also creating dependency on interactional and social dynamics that condition how evidence is conferred. Judgement in assessed interviews is therefore selective, tacitly interpreted, and relationally stabilised.

Interview questions pre-authorise what counts as a legitimate answer

Interview questions emerged as active design devices that shape what is recognisable as a legitimate answer before students begin to speak. Rather than operating as neutral prompts, questions were understood, by both staff and students, as framing the kinds of accounts that could count within the assessment encounter. Legitimacy was not evaluated solely through reference to formal marking criteria, but was pre-authorised through the structure and focus of the questions themselves. Staff accounts made this design logic explicit. One staff member reflected that while students were engaged in substantial work and learning across the programme, the interview reframed what was assessable by narrowing attention to particular kinds of reflective articulation: '…in the interview they are asked very specifically, "What did you learn from that process?"' (ST1). Judgement is positioned as question-led

rather than criterion-led, with the interview functioning to translate experience into a specific evaluative form. Another staff member similarly noted that the framing effects of the assessment were often invisible to students, despite being structurally embedded. This framing was described not as incidental, but as intentionally organised through the questions themselves, with one staff member noting that 'the taxonomy is there, embedded into the questions… their responses are framed in a way that is focused by the taxonomy' (ST2).

From the student perspective, these pre-authorisations were most visible at moments of uncertainty or misalignment, particularly early in the assessment trajectory. One student described struggling to locate what was being asked of them, not in terms of content knowledge but in terms of legitimacy: 'I didn't know what counted as reflection or what you wanted from me' (SP10). This uncertainty was not expressed as a lack of preparation or effort, but as difficulty interpreting what the question was inviting, and excluding. The same student went on to reflect that this experience was not isolated to a single assessment, but characteristic of a particular category of interview questions: 'I've never thought I've done very well on those questions, you know, "give an example of a time you learned something" or "reflect on something"' (SP10). These accounts suggest that the challenge lay not in having experiences to draw on, but in discerning how those experiences needed to be framed to meet the implicit expectations embedded in the questions. Reflection functioned less as an open-ended invitation and more as a regulated form of self-accounting, with legitimacy contingent on alignment with a tacit answer genre. As one staff member noted: 'They may not all have noticed this… but the structure of the assessment means their responses are framed' (ST2).

Importantly, this pre-authorisation was not always fully transparent even to assessors themselves. One staff member captured this ambiguity when reflecting on marking discussions: 'What's the important thing? What are we marking here?' (ST4). This question points to the tacit and negotiated nature of the judgement system underpinning interview assessment. While questions structured responses in advance, the standards against which those responses were evaluated were not always fully explicit or stabilised, reinforcing the reliance on question framing as a practical mechanism for guiding judgement.

Rather than capturing lived experience across the assessment encounter, this theme addresses the design assumptions embedded in interview questions that shape what counts as a legitimate answer before students speak. These assumptions are often tacit and therefore rarely articulated directly by participants, becoming visible primarily at moments of misalignment, uncertainty, or explicit design reflection. As students become familiar with the interview format, these assumptions recede from explicit attention and are taken up through preparation practices and interactional adjustment.

Answer genres are produced interactionally, not delivered whole

Interview answers were not treated by participants as pre-formed performances that students simply brought into the assessment encounter. Instead, what came to count as a 'good' answer frequently emerged through the interaction itself, shaped by

questions, follow-up prompts, conversational flow, and assessor uptake. This theme highlights assessed interviews as sites where answers are produced in real time, rather than merely revealed.

Students often described moments where examples or connections only became available once the conversation was underway. For some, the question itself acted as a trigger that surfaced experiences they had not anticipated drawing on. One participant reflected that when asked about teamwork, they suddenly recalled managing a crisis at work, noting that 'the conversation helped me find examples I'd forgotten' (SP1). The answer did not exist in advance as a rehearsed narrative, but was activated through the interaction. Several students described their thinking becoming clearer through the act of speaking. Rather than spoken answers simply translating prior cognition into oral form, articulation itself appeared to be a means through which coherence was achieved. One participant noted that although they were initially nervous, 'once I started talking, I realised I could explain my ideas more clearly out loud than I expected' (SP8). Another explicitly contrasted spoken and written modes, explaining that while writing required pre-structuring, 'my thinking evolves as I go' when speaking face to face (SP4). In these accounts, judgement-relevant qualities such as clarity and organisation were not fully present at the outset of an answer, but emerged as the answer unfolded.

Answers also became legitimate through interactional recognition by others. In some cases, competence was retrospectively named into being by an assessor. One student recalled an interviewer responding to their account by saying, 'That's leadership, you just hadn't named it yet,' a moment the student described as consequential (SP2). The issue was not a lack of experience, but a lack of recognised framing. Legitimacy was conferred through the interaction, transforming a personal account into an assessable demonstration of competence. A similar dynamic was evident in coaching contexts, where students described bringing 'messy notes' into a session and having a coach respond, 'Let's find the skill in this' (SP9). Interaction functioned as a mechanism for translating experience into recognised assessment currency.

Participants frequently contrasted this dialogic process with more checklist-oriented understandings of assessment. Several emphasised that the interview did not feel like a matter of ticking off criteria, but rather like a conversation in which meaning developed collaboratively. One student noted, 'I don't feel like it was a ticking a sheet for the tutor… it was more of a dialogue and flow' (SP5), while another described the interview as 'surprisingly natural… like a conversation about the professional version of me' (SP2). These accounts do not suggest an absence of judgement, but rather a different mode of judgement: one enacted through conversational alignment and shared sense-making rather than through discrete, observable outputs.

Producing answers interactionally was described as cognitively and relationally demanding. Students spoke about managing multiple simultaneous tasks: listening, thinking, monitoring delivery, and recalling examples, while responding in real time. One participant described 'juggling a lot at once… listening, thinking, making eye contact, trying to remember my examples without looking like I was remembering them' (SP3). Such accounts foreground the bandwidth constraints under which interactional judgement is enacted, underscoring that answers are assembled under pressure rather than calmly delivered as finished products.

Staff accounts reveal tensions between answer content and interactional performance. One staff member reflected on a student who 'gave really good answers… but in many ways I suppose it was a bad interview' (ST4), while also describing prolonged deliberation between markers in cases where 'the examples were good, but they weren't necessarily delivered confidently' (ST4). These reflections indicate that interactional production does not eliminate ambiguity from judgement; instead, it relocates it. Assessors are required to interpret how emerging answers should be weighed, particularly when content and delivery do not align neatly. Merit is not simply demonstrated but co-produced. Answers take shape through interaction, become legitimate through recognition, and are judged within the constraints of conversational flow and temporal pressure.

Preparation is oriented towards recognisability rather than self-expression

Participants described preparation for assessed interviews not simply as acquiring content or improving delivery, but as a process of anticipating how their responses would be judged and calibrating themselves accordingly. Preparation was oriented towards producing answers that would be recognisable as legitimate within the interview context, often framed as a form of risk management rather than as a pursuit of authenticity or self-expression. Preparation extended the judgement encounter beyond the interview itself, shaping how students selected and rehearsed what they would say in advance.

Several participants articulated preparation as a protective strategy designed to minimise the risk of misrecognition under time pressure. One student reflected, 'I prepared a lot. Probably too much… I didn't want the "real me", the one who gets flustered, to show up unprepared. It felt like too much of a risk' (SP3). Preparation functioned not as mastery but as containment, allowing students to manage how they might be seen in a high-stakes, spoken assessment. This framing complicates deficit interpretations of over-preparation, positioning it instead as a rational response to anticipated judgement.

Preparation was also described as learning the form of a legitimate answer, rather than simply drawing on experience. Participants noted that experience alone was insufficient unless it could be framed in ways that aligned with what interview questions implicitly demanded. As one student explained, 'I also knew this interview wasn't just about telling a story. It wasn't enough to say, "This happened to me." I had to show why that story mattered, what it demonstrated, what it said about how I think or what I've learned' (SP12). Another reflected on this shift retrospectively: 'Now I can say, "Here's what I learned from that situation, and here's how I've applied it since"' (SP10). These accounts suggest that preparation involved internalising recognisable reflective templates, enabling students to translate experience into assessable accounts.

The necessity of such preparation became particularly visible when students described moments where experience failed to become legible. One participant recounted freezing during an interview despite having relevant experience: 'I froze because I didn't know the words… I had loads of experience but I didn't know how to talk about it without rambling' (SP9). The difficulty here was not a lack of

experience, but a lack of access to the language and framing that would allow that experience to count. Preparation thus emerged as a means of bridging the gap between lived experience and institutional recognisability.

Participants also described preparation as an iterative process shaped over time through repeated exposure to interview formats. Earlier, lower-stakes encounters functioned as rehearsals that clarified expectations and informed later preparation. As one student noted, 'The [Level Four] interview really prepped me well. It really prepared me for the second [Level Five] interview' (SP7). Another commented on the reassurance provided by advance knowledge of questions: 'The Level Four one was okay, I liked having the questions in advance. I prepared answers and I think I did alright' (SP11). These accounts suggest that preparation was not merely individual choice, but was scaffolded through assessment design, gradually normalising particular preparation practices.

Preparation also emerged as a means of managing tensions between recognisability and authenticity. Rather than presenting preparation as inauthentic performance, participants described selecting which version of themselves felt safest to present under assessment conditions. One student reflected, 'It's like I'm putting on this "professional" version of me, someone I'm still getting to know, or maybe not quite ready to be yet' (SP3). Another described the moral difficulty of resisting perceived expectations: 'Resisting the urge to say what I thought you wanted to hear, that was hard' (SP2). Preparation entailed ethical and identity work, as students navigated how to remain sincere while still aligning with anticipated judgement norms.

Discussion

Design as pre-authorisation

Assessed interviews emerged as assessment formats that organise what can count as evidence in advance of judgement. Interviews therefore work through pre-authorisation: the prospective structuring of evidential possibilities by design. In interview assessment, pre-authorisation is analytically distinct from evaluation. Evaluation concerns how assessors interpret responses once produced. Pre-authorisation concerns how assessment architecture positions assessors to recognise some forms of response as relevant and credible in the first place, a distinction that extends judgement-oriented accounts of assessment practice (Shay 2005). In interview assessment, this organisation is enacted primarily through questions, which frame not only what is asked but how answers must be structured in order to become legible as assessable evidence within the encounter.

The effects of pre-authorisation become decisive under the temporal and interactional constraints of interview assessment, where judgement must be enacted in real time and evidential exploration is necessarily selective. Responses that align closely with the framing of the question are more readily stabilised as evidence, while those requiring translation, reframing, or alternative modes of demonstration are less easily taken up. In assessor-led, time-bounded interactional contexts such as interviews, criteria rarely function as moment-by-moment guides within the interaction. Instead, assessors rely on format-relevant cues that signal adequacy

within the compressed encounter. Pre-authorisation tends to channel assessor discretion, producing patterned recognition shaped by design choices about questions, timing, and interactional form.

Understanding interview assessment through the lens of pre-authorisation clarifies how familiar claims about fairness, consistency, and transparency are typically framed, and why they are often only partial. In assessment literature, fairness can be framed in terms of equivalent questioning or unbiased judgement, consistency in terms of agreement between assessors, and transparency in terms of the clarity of criteria or expectations in advance (e.g. Nallaya et al. 2024; Ragupathi and Lee 2020). This paper suggests that such claims cannot be interpreted solely at the point of judgement. In interview assessment, fairness and consistency are already shaped upstream, in how questions delimit what can count as relevant, credible, or assessable within the encounter. In interview assessment, transparency cannot be reduced to the disclosure of criteria alone when much of what governs recognition emerges interactionally and contingently under time constraint. This relocates these concepts from assessor behaviour to assessment architecture.

Compressed judgement

Interview assessment operates through a distinct judgement logic, here termed compressed judgement. Judgement is not merely constrained by time and interactional pressure but takes shape through them. Because the interview is a bounded and irreversible encounter, evaluation shifts away from the accumulation of evidence and towards the production of accounts that are decision-ready in the moment. Under compression, evidential adequacy is oriented not towards completeness or representativeness, but towards interactional sufficiency: whether an account can be recognised, stabilised, and taken up as evidence within the unfolding encounter. This analysis distinguishes compressed judgement from deliberative or accumulative forms of assessment. Rather than differing by degree of rigour, the distinction lies in the conditions under which evidential adequacy is established: a concern that judgement scholarship has typically examined in relation to more extended evaluative processes (Sadler 2009).

Common claims that oral assessment affords deeper or more diagnostic access to understanding are not rejected here, but reinterpreted as effects of interactional recognisability under conditions of compression, rather than as properties of the evidence. Under compressed judgement, interaction is not simply the medium of assessment but the evaluative process itself. Compressed judgement therefore tends to privilege forms of merit that can be recognised in action, resonating with accounts of assessment judgement that conceptualise standards as enacted in practice (Ajjawi, Bearman, and Boud 2021). Judgement advances through cycles of prompting, response, uptake, and adjustment, in which evidential status is continuously negotiated in real time. What succeeds under these conditions is not completeness, but recoverability: the capacity of an account to survive compression through clarification and redirection within the interaction. These interactional adjustments are the means by which evidence remains usable under conditions of immediacy.

Understanding interview assessment as operating through compressed judgement reframes many enduring tensions not as problems to be resolved, but as structural conditions of evaluation. Difficulties reconciling depth with responsiveness, completeness with time pressure, or deliberation with irreversibility follow from a judgement logic oriented towards decision-readiness within a bounded and unfolding encounter. These tensions are therefore predictable consequences of evaluating evidential adequacy under interactional and temporal constraint, rather than artefacts of poor practice or assessor inconsistency.

This paper extends judgement-theoretical scholarship not by refining how standards are interpreted, but by reconfiguring where judgement is understood to operate. Work on interpretive judgement, enacted standards, and performative standards has shown that assessment decisions cannot be reduced to the application of criteria and are socially and institutionally patterned (Sadler 2009; Shay 2005). The present analysis shifts attention from how standards are interpreted to how evidential adequacy becomes decision-ready under conditions of temporal irreversibility and interactional constraint. Compressed judgement is not a general theory of assessment judgement, but a judgement logic associated with formats such as interview-based assessment, in which evaluation must be enacted within a bounded and unfolding encounter. It foregrounds the role of time, interaction, and recognisability in shaping what can count as evidence, dimensions that remain under-specified in existing accounts of judgement.

Contribution, scope, and limits

This study contributes a judgement-theoretical account of interview-based assessment by conceptualising interviews as formats that organise recognition under conditions of temporal compression and interaction. It reframes recurring features of interview-based assessment that are often treated as problems of practice or judgement: why some forms of evidence consistently count more than others at the point of evaluation, why patterned judgements emerge despite underspecified criteria, and why efforts to value depth, authenticity, or breadth are difficult to realise in real time. By locating these patterns in assessment design and judgement conditions rather than in individual performance or assessor behaviour, the paper offers a clearer basis for analysing how evidential adequacy becomes recognisable in oral assessment formats. This supports examining spoken assessment as a structured judgement system rather than as variable interactional performance.

The contribution is conceptual rather than representational. While grounded in a single programme, the analysis addresses judgement conditions characteristic of assessor-led, time-bounded interview assessments more broadly. Its value lies in offering an analytic lens for examining how evidential adequacy becomes recognisable in similar formats, rather than in making claims about practices or outcomes across contexts. Although the programme from which this analysis arises employs criterion-referenced ipsative assessment, the analysis does not depend on ipsative comparison as a judgement mechanism. The limits of the study are constitutive of this focus. It does not make normative claims about the fairness or desirability of

interviews, offer prescriptive guidance for assessment design, or evaluate whose experience is more or less likely to be recognised under conditions of compressed judgement, where recognisability is unlikely to be evenly distributed. Examining how compressed judgement intersects with social positioning and institutional norms therefore remains a priority for future research.

Work on oral and viva-style assessment has long acknowledged that interactional structure and examiner questioning shape what candidates are able to demonstrate. Joughin (1998) conceptualises oral assessment as inherently interactional, showing how design features such as question framing, sequencing, and examiner follow-up shape the forms of response that can emerge. By contrast, structured viva research in health professions education has often approached interaction primarily as a source of variability to be managed in the interests of reliability (e.g. van der Vleuten and Schuwirth 2005). The present analysis treats interaction not as variance or design feature, but as part of the judgement logic of assessment, opening up further lines of inquiry into how different oral assessment designs pre-authorise particular evidential pathways, how assessors navigate evidential sufficiency under time constraint, and how assessment formats shape the conditions under which evidence becomes usable in judgement.

Conclusion

Interview-based assessment operates as a judgement format in which evidential sufficiency is established under conditions of temporal and interactional compression. Where decisions must be made within a bounded and irreversible encounter, judgement cannot rely on the accumulation or later inspection of evidence. Instead, evidential adequacy is produced in real time through questioning and interaction. What becomes judgeable is therefore not what is extensive or representative, but what can be rendered intelligible and decision-ready within the encounter.

Seen in these terms, interview assessment is organised less around the evaluation of pre-existing evidence than around the conditions that make evidence recognisable at the point of judgement. Question design delimits the forms of response that can count in advance, interaction provides the means through which accounts are clarified or redirected, and preparation becomes oriented towards producing contributions that can survive compression. These elements function together as a judgement logic oriented towards immediacy.

Understanding interview assessment in this way reorients how persistent features of assessor-led oral formats are interpreted. Patterned judgements, the recurrent privileging of certain kinds of evidence, and the difficulty of realising aspirations towards depth or breadth in real time are not artefacts of inconsistent marking. They follow from the structural conditions under which judgement is enacted. Attention therefore shifts from individual performance or assessor behaviour to the architectural features of assessment that organise recognition under constraint. Future work might extend this analysis by examining how different questioning architectures reshape what becomes recognisable as evidence and how assessors collectively calibrate interactional sufficiency.

Declaration and ethics statement

Certificate of Research Ethics Approval from Birmingham Newman University, dated 11 March 2025. Application number: S2024/012. PDF of Certificate provided.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Emma J. Folwell http://orcid.org/0000-0001-7792-2650

Data availability statement

Data available on request due to privacy/ethical restrictions.

References

Ajjawi, R., M. Bearman, and D. Boud. 2021. "Performing Standards: A Critical Perspective on the Contemporary Use of Standards in Assessment." *Teaching in Higher Education* 26 (5): 728–741. https://doi.org/10.1080/13562517.2019.1678579.

Bloxham, S., P. Boyd, and S. Orr. 2011. "Mark my Words: The Role of Assessment Criteria in UK Higher Education Grading Practices." *Studies in Higher Education* 36 (6): 655–670. https://doi.org/10.1080/03075071003777716.

Braun, V., and V. Clarke. 2022. *Thematic Analysis: A Practical Guide*. Los Angeles, CA: SAGE.

Crecelius, A. R., L. R. DeRuisseau, and J. Brandauer. 2021. "Cumulative Oral Examinations in Undergraduate Human Physiology: Process, Student Perceptions, and Outcomes." *Advances in Physiology Education* 45 (3): 518–525. https://doi.org/10.1152/advan.00028.2021.

Dickson, H., J. Harvey, and N. J. Blackwood. 2019. "Feedback, Feedforward: Evaluating the Effectiveness of an Oral Peer Review Exercise Amongst Postgraduate Students." *Assessment & Evaluation in Higher Education* 44 (5): 692–704. https://doi.org/10.1080/026029s38.2018.1528341.

Ecclestone, K. 2001. ""I Know a 2:1 When I See It": Understanding Criteria for Degree Classifications in Franchised University Programmes." *Journal of Further and Higher Education* 25 (3): 301–313. https://doi.org/10.1080/03098770126527.

Fischer, L. E., M. Snyder, S. A. Sullivan, E. F. Foley, and J. A. Greenberg. 2016. "Evaluating the Effectiveness of a Mock Oral Educational Program." *The Journal of Surgical Research* 205 (2): 305–311. https://doi.org/10.1016/j.jss.2016.06.088.

Harsy, A. 2024. "Using Oral Exams to Assess Conceptual Understanding in Proof Writing Courses." In *Effective Alternative Assessment Practices in Higher Education*, edited by A. D. Tomlin and C. M. Nowik, 177–194. York: Information Age Publishing.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. https://doi.org/10.3102/003465430298487.

Huxham, M., F. Campbell, and J. Westwood. 2012. "Oral versus Written Assessments: A Test of Student Performance and Attitudes." *Assessment & Evaluation in Higher Education* 37 (1): 125–136. https://doi.org/10.1080/02602938.2010.515012.

Joughin, G. 1998. "Dimensions of Oral Assessment." *Assessment & Evaluation in Higher Education* 23 (4): 367–378. https://doi.org/10.1080/0260293980230404.

Laurin-Barantke, L., Hoyer J., Fehm L., Knappe, S. 2016. Oral but Not Written Test Anxiety is Related to Social Anxiety.*World Journal of Psychiatry*.6;6(3):351–357. https://doi.org/10.5498/wjp.v6.i3.351.

McBain, B., A. Drew, C. James, L. Phelan, K. M. Harris, and J. Archer. 2016. "Student Experience of Oral Communication Assessment Tasks Online from a Multi-Disciplinary Trial." *Education + Training* 58 (2): 134–149. https://doi.org/10.1108/ET-10-2014-0124.

Moorhouse, B. L., M. A. Yeo, and Y. Wan. 2023. "Generative AI Tools and Assessment: Guidelines of the World's Top-Ranking Universities." *Computers and Education Open* 5: 100151. https://doi.org/10.1016/j.caeo.2023.100151.

Nallaya, S., S. Gentili, S. Weeks, and K. Baldock. 2024. "The Validity, Reliability, Academic Integrity and Integration of Oral Assessments in Higher Education: A Systematic Review." *Issues in Educational Research* 34 (2): 629–646.

Nelson, M. A. 2010. "Oral Assessments: Improving Retention, Grades, and Understanding." *Primus* 21 (1): 47–61. https://doi.org/10.1080/10511970902869176.

Oliveira, A. W., A. O. Brown, M. L. Carroll, E. Blenkarn, B. Austin, and T. Bretzlaff. 2021. "Developing Undergraduate Student Oral Science Communication through Video Reflection." *International Journal of Science Education, Part B* 11 (2): 143–154. https://doi.org/10.1080/21548455.2021.1907630.

Ragupathi, K., and A. Lee. 2020. "Beyond Fairness and Consistency in Grading: The Role of Rubrics in Higher Education." In *Diversity and Inclusion in Global Higher Education*, edited by C. S. Sanger and N. W. Gleason, 73–95. Singapore: Palgrave Macmillan.

Rust, C, M. Price, and B. O'donovan. 2003. "Improving Students' Learning by Developing Their Understanding of Assessment Criteria and Processes." *Assessment & Evaluation in Higher Education* 28 (2): 147–164. https://doi.org/10.1080/02602930301671.

Sadler, D. R. 2009. "Indeterminacy in the Use of Preset Criteria for Assessment and Grading." *Assessment & Evaluation in Higher Education* 34 (2): 159–179. https://doi.org/10.1080/02602930801956059.

Sato, B. K., C. F. Hill, and S. M. Lo. 2019. "Testing the Test: Are Exams Measuring Understanding?" *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology* 47 (3): 296–302. https://doi.org/10.1002/bmb.21231.

Shay, S. 2005. "The Assessment of Complex Tasks: A Double Reading." *Studies in Higher Education* 30 (6): 663–679. https://doi.org/10.1080/03075070500339988.

Sox, Colin M., Rebecca Tenney-Soeiro, Linda O. Lewin, Jeanine Ronan, Mary Brown, Marta King, Rachel Thompson, et al. 2018. "Efficacy of a Web-Based Oral Case Presentation Instruction Module: Multicenter Randomized Controlled Trial." *Academic Pediatrics* 18 (5): 535–541. https://doi.org/10.1016/j.acap.2017.12.010.

Stephenson, Z., N. Johnson-Glauch, and S. Cruchley. 2025. "Interventions and Facilitators of Oral Assessment Performance in Higher Education: A Systematic Review." *Assessment & Evaluation in Higher Education* 50 (7): 1140–1153. https://doi.org/10.1080/02602938.2025.2504621.

Sullivan, M., A. Kelly, and P. McLaughlan. 2023. "ChatGPT in Higher Education: Considerations for Academic Integrity and Student Learning." *Journal of Applied Learning and Teaching* 6 (1): 31–40.

van der Vleuten, C. P., and L. W. Schuwirth. 2005. "Assessing Professional Competence: From Methods to Programmes." *Medical Education* 39 (3): 309–317. https://doi.org/10.1111/j.1365-2929.2005.02094.x.

Van Ginkel, S., J. Gulikers, H. Biemans, and M. Mulder. 2015. "Principles for Developing Oral Presentation Competence: A Systematic Review." *Educational Research Review* 14: 62–80.

Van Ginkel, S., J. Gulikers, H. Biemans, and M. Mulder. 2017. "The Impact of Feedback Source on Developing Oral Presentation Competence." *Assessment & Evaluation in Higher Education* 42 (6): 953–966.

Appendix

The following guides were used to structure semi-structured interviews. Questions were asked flexibly and adapted in response to participants' accounts. Not all prompts were asked of all participants, and additional probes were used where clarification or elaboration was required.

Staff interview guide

Context

- Thinking about the modules, what stands out most to you about how they work in practice?
- Can you describe a moment or example that captures what these modules are trying to achieve?

Assessment

- When designing or assessing these tasks, what kinds of learning or ability are you hoping to make visible?
- How do students demonstrate that learning?
- How do you see interview assessments functioning?
- Are there tensions between academic and professional expectations?

Reflection and feedback

- How do reflection and feedback operate within these modules?
- What helps students recognise their development?
- Have you noticed shifts in how students articulate their learning over time?

Knowledge and translation

- How do students connect or translate subject learning into applied or professional contexts?
- Can you recall examples where this translation was successful or challenging?

Relationships

- How do relationships (e.g., between staff and students) shape assessment and feedback practices?
- Are there moments where issues of fairness or inclusion become salient in assessment?

Programme development

- How has your approach to employability or assessment evolved since working on this programme?
- How do assessment structures (e.g., portfolios, interviews, coaching) support or constrain learning?
- What would you like to develop or change in how assessment operates?

Closing

- What do you think these modules ultimately enable students to take forward?

Student interview guide

Context

- When you think back to the portfolios and interviews, what stands out most for you?
- Were there particular moments that felt important or memorable?
- How did the different interview assessments (e.g., Level Four, Level Five) compare?

Experience of assessment

- How did it feel to talk about your work and learning in the interviews?
- Was that experience different from the portfolio assignments?
- What did "doing well" in the interview mean to you?
- How did you decide what to include in your answers?

Preparation and feedback

- How did you prepare for the interviews?
- How did feedback influence how you approached later interviews?
- Did your preparation change over time?

Expectations

- How clear were the expectations for the interview assessments?
- How did you understand what was being assessed?
- Were there moments when you felt unsure about what counted as a strong answer?

Knowledge and translation

- What kinds of experiences or learning did you draw on in the interviews?
- Did you find yourself explaining or adapting your ideas differently in the interview format?
- Were there things that felt easier or harder to express in an interview compared to written work?

Development over time

- Do you think your approach to the interviews changed as you moved through the degree?
- Looking back, what have you learned about presenting or explaining your ideas?
- If you could go back to your first interview, what would you do differently?

Closing

- Overall, what do you think these interview assessments were asking you to demonstrate?
- Is there anything about the interview format that particularly helped or challenged you?