Assessment in Education: Principles, Policy & Practice

ISSN: 0969-594X (Print) 1465-329X (Online) Journal homepage: www.tandfonline.com/journals/caie20

Less versus, more &: polarities as gradations in research on student involvement in assessment

Ernesto Panadero

To cite this article: Ernesto Panadero (21 Oct 2026): Less versus, more &: polarities as gradations in research on student involvement in assessment, Assessment in Education: Principles, Policy & Practice, DOI: 10.1080/0969594X.2026.2752430

To link to this article: https://doi.org/10.1080/0969594X.2026.2752430

Published online: 21 Oct 2026.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=caie20

ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE
https://doi.org/10.1080/0969594X.2026.2752430

Routledge
Taylor & Francis Group

Check for updates

Less versus, more &: polarities as gradations in research on student involvement in assessment

Ernesto Panadero [a,b,c]

[a]Centre for Assessment Research, Policy and Practice in Education (CARPE), Dublin City University, Dublin, Ireland; [b]Education Regulated Learning and Assessment (ERLA group), Facultad de Educación y Deporte, Universidad de Deusto, Bilbao, Spain; [c]Ikerbasque, Bilbao, Spain

ABSTRACT

Research on student involvement in assessment has matured over four decades. Yet much of the field organises itself around polarities: summative versus formative, psychology versus pedagogy, accuracy versus agency, tradition versus innovation. I argue that these polarities are more productive as gradations than as oppositions. Real disagreements remain, but they are better argued on their merits than organised into camps. Three claims follow. First, the oppositions are looser in practice: researchers associated with one pole routinely work on questions from the other. Second, standing on the shoulders of giants means engaging seriously with prior work, foundational and recent, rather than claiming novelty by passing over it. Third, the field shares methodological problems: weak designs, non-uniform measurement, unclear specification of interventions, and an underdeveloped account of how students engage with assessment. I propose an integration test for these debates and recommendations for authors.

ARTICLE HISTORY
Received 29 May 2026
Accepted 16 October 2026

KEYWORDS
Higher education; research methodology; reporting recommendations; self-assessment; student involvement in assessment; peer assessment

This is a short paper. It is also a personal one. I have spent the last twenty years working in the field of student involvement in assessment, including self-assessment and peer assessment, and a substantial part of that work has consisted in reviewing what others have done. By student involvement in assessment, I mean practices in which students judge their own work, judge their peers' work, or help set assessment criteria. Meta-analyses and systematic reviews are the form that reviewing has often taken. I work this way because I believe that the best place to stand when developing a research programme is on a careful synthesis of what is already known. The field has produced a great deal that is worth knowing, and the periodic exercise of stepping back to take stock has been the basis on which my own thinking has developed.

From that vantage point, I want to put forward a position on how the field organises itself. By a polarity, I mean a pair of approaches presented as mutually exclusive, such as summative versus formative assessment. By a gradation, I mean a dimension along which a study can take more or less of each approach. Commitment to each approach is a matter of degree, not of kind, and a strong commitment to one does not rule out a strong

CONTACT Ernesto Panadero ernesto.research@gmail.com Carpe, Dublin City University, St Patrick's Campus, Drumcondra, Dublin 9, Ireland

© 2026 Informa UK Limited, trading as Taylor & Francis Group

commitment to the other. Treating a polarity as a gradation is therefore not a call to compromise at a midpoint. The position is that the polarities around which the field sometimes structures its debates are less productive when treated as binary oppositions and more productive when treated as gradations along which different research questions and theoretical and pedagogical priorities sit. This is a small claim, but I believe it has consequences for how the field grows. Two recent papers examine one of these polarities in depth, the relation between accuracy and agency in self-assessment (Panadero & Brown, 2026a, 2026b). This note does not repeat their arguments. It steps back to student involvement in assessment as a whole, including peer assessment and feedback. It asks how the field organises its debates and where its most useful work lies. The paper develops the position in three movements: first on the polarities themselves, second on the importance of standing on the shoulders of giants, and third on the methodological problems that the field would do well to address through integrative rather than polarising work.

Less versus, more &

Educational assessment has organised much of its life around several polarities. Summative versus formative. Psychology versus pedagogy. Large-scale and testing-focused work versus classroom-focused practice. Tradition versus innovation. These polarities are not invented; they correspond to real differences in research traditions, methodological commitments, and pedagogical priorities, and the differences sometimes matter; but the polarities also have a way of organising debate that is less helpful than it looks. They suggest that researchers must choose a side, that the sides are mutually exclusive, and that the work of the field consists in the contest between them. None of this is true, especially if our aim is to move the field forward.

In the actual practice of research, the polarities are more like gradations. Studies have more or less of a summative orientation, more or less of a formative purpose, more or less psychological focus, more or less pedagogical engagement. The same study can sit at different points along several of these dimensions at once. Meta-analyses on the effects of self-assessment on academic performance (e.g. Yan et al., 2022) are psychological in their analytic apparatus and pedagogical in their motivation; they speak to formative purposes through summative outcome measures. A study on peer assessment training (Sluijsmans et al., 2002) addresses an apparently technical question (how do we teach students to assess?) that is at the same time deeply pedagogical (what kind of learners do we want to develop?). Treating these studies as falling on one side or the other of a binary loses what they actually contribute.

There is a more revealing observation than this, however, that I think the field would benefit from acknowledging. Researchers who are most closely associated with one pole of these polarities frequently produce work that engages questions associated with the other. This openness is to their credit. To pick three examples that I find instructive: Sadler's account of formative assessment is foundational to the formative tradition (Sadler, 1989). Yet much of his later work aims to make grading trustworthy, from grade integrity to the calibration of academic standards (Sadler, 2005, 2009, 2013). Black and Wiliam's (1998) review is equally foundational. Black and colleagues later studied whether teachers' summative assessments can produce dependable results and also

enhance classroom learning (Black et al., 2011). Similarly, H. L. Andrade (2019) has argued for less research on consistency and summative self-assessment, and more on the mechanisms of formative self-assessment. In the same review, she calls for research on how accuracy in formative self-assessment relates to students' subsequent revisions and learning (H. L. Andrade, 2019, p. 7). The polarisation between a formative-pedagogical tradition and a summative-psychometric one does not survive close contact with what these scholars actually study. Even researchers whose explicit theoretical commitments lie squarely on the formative-pedagogical side find the questions of accuracy, calibration, and judgement compelling enough to research with empirical seriousness.

This is not, in my reading, a contradiction in the work of those scholars. It is a sign that the field as it is actually practised does not respect the polarities some scholars try to identify. The same pattern appears in the current debate on agency in self-assessment, examined in detail in Panadero and Brown (2026a, 2026b). The questions are too interesting and the practical stakes too real for any researcher in this area to fully inhabit one pole of the supposed dichotomy. What follows from this is straightforward. The narration of the field as a contest between traditions does the field a disservice. The narration that takes the field seriously is one in which we acknowledge that there is more and less of various commitments operating in any piece of work, and that these commitments coexist productively rather than substituting for each other.

In sum: less versus, more &. Treating the polarities as gradations does not dissolve every tension between them. There are genuine disagreements about the aims of education, the meaning of agency, what counts as evidence, and whose interests assessment serves. A psychometric view of assessment, which is inferential, and a qualitative view, which is not, rest on different epistemologies. That a single study combines formative purposes and summative measures shows that the poles are not mutually exclusive. It does not show that these disagreements are merely rhetorical. My claim is narrower. These disagreements are better argued on their merits, within shared empirical work, than organised into camps. I call this the *integration test*: whether a position proposes to integrate the poles or to replace one with the other. The difference between treating the polarities as gradations and treating them as binaries matters, and the work of the present generation is not improved by the rhetorical move of replacing one pole with another. But the polarities, taken at their worst, are not where the most serious problems of the field lie. Those problems concern what happens when the foundational architecture of the field is set aside in the name of innovation, and what happens when the empirical work that follows is conducted without the methodological care its claims require. The next two sections take each of these in turn.

Standing on the shoulders of giants

If the field is built across decades through gradations rather than oppositions, then the work of the present generation should be in continuity with the work of those who came before. This sounds obvious, but it has implications that are sometimes resisted. The implication is that the empirical and conceptual contributions of foundational figures cannot be set aside in the name of innovation without losing the architecture on which contemporary work depends. Boud and Falchikov's (1989) critical analysis of self-assessment research, Falchikov and Boud's (1989) meta-analysis of student-teacher

mark agreement, Sadler's (1989) account of formative assessment and expert evaluative judgement, Black and Wiliam's (1998) consolidation of formative assessment as a research programme, Andrade and Valtcheva's (2009) framing of self-assessment as a metacognitive process tied to motivation, self-efficacy, and self-regulated learning, Brown and Harris's (2013) synthesis of K-12 self-assessment research, all of these are part of the foundation. We do not have a field without them. Panadero and Brown (2026b) set out what this foundation contains for self-assessment and why it still matters. The same holds for peer assessment, where early syntheses of reliability, validity, and effects remain reference points (Falchikov & Goldfinch, 2000; Topping, 1998).

My own contribution to the field has been to spend a substantial portion of my time photographing the work that others do, trying to understand it, and synthesising it for the field as a whole. This is what reviews and meta-analyses are. They are, or at least should aim to be, exercises in taking stock of what colleagues across institutions, traditions, and decades have produced, and in articulating what that production tells us collectively, and not in trying to bend or bias the conclusions to fit the researcher's narrative. The point of this work has not been to defend a particular methodological position. The point has been to take seriously what the field has produced and to articulate, on the basis of that production, what we know and what we do not.

This way of working has consequences for how I think about the field's polarities. When you spend years reading what colleagues across the supposed dividing lines of the field have produced, the dividing lines themselves come to look less convincing. A meta-analysis does not respect the boundary between the psychometric and the pedagogical, the summative and the formative, the traditional and the innovative. It cannot afford to. The studies it integrates do not respect that boundary either. The result is a kind of forced humility about the field. Whatever pole one's own theoretical sympathies might be drawn to, the empirical record is built by everyone, and the standing one has to claim is the standing of having engaged seriously with the whole of it. This kind of engagement carries its own demands. Reading across traditions is not, in itself, a substitute for methodological care, and what counts as serious engagement with the empirical record is not a matter of personal preference. When a synthesis sets aside studies with weak designs, ambiguous measures, or insufficiently specified interventions, it is not enforcing a disciplinary aesthetic. It is protecting the conclusions the field will draw from the work that remains. Methodological exigency in this area is a precondition for the trustworthiness of what we say collectively, regardless of the theoretical framework under which any given study was conducted.

Standing on the shoulders of giants is a phrase that gets used loosely. I take it seriously here. The figures whose work has shaped the field are owed the kind of reading that takes their contributions in their own terms, with the empirical care they themselves brought to the questions they posed. They are not owed deference, but they are owed engagement. The recent literature has sometimes risked moving too quickly past this engagement in pursuit of new framings, and the cost of that movement is a thinner field, not a richer one. The richer field is one in which the foundational architecture is preserved and the new questions are posed in continuity with it. A symptom worth noting is the increasing use of the phrase 'to our knowledge' as a warrant for novelty in places where a more careful reading of prior work would have shown the claim to be less original than presented. The phrase has acquired a second function: not signalling the limits of what

authors have read, but justifying not having read further. The point is not to abandon the phrase. It is to ensure that, when it appears, it follows from a serious attempt to read the prior work rather than substituting for it. This is what I mean by tradition and innovation rather than tradition versus innovation. And the obligation of careful reading is not reserved for the foundational figures alone. It applies, as a matter of basic scholarly practice, to any prior work the present argument depends on, whatever the seniority of those who produced it. The fact that a body of recent research can be passed over with a single phrase does not mean that it has been engaged with, and engagement, not dismissal, is what the cumulative development of the field requires.

Where the field needs to do better work

If the polarities are looser than they appear and the foundational and previous work is owed serious engagement, then the most useful work the field can do at present is the work of addressing its real methodological problems. There are several of these, and they cut across the supposed traditions. Naming them clearly seems to me the most concrete way of articulating what an integrative agenda for the field looks like. This is not a claim that the empirical programme has stalled. Some of its recent work is among the most methodologically innovative in assessment research (Panadero & Brown, 2026b). The problems below concern much of the published literature rather than its best work. My own studies are part of that literature, and some of them fall short of the recommendations I make below.

The first problem concerns the quality of research designs in the field of student involvement in assessment. A substantial part of the empirical literature relies on weak designs: studies without proper control conditions, studies with intervention specifications too thin to be replicable, studies whose dependent variables are not adequately operationalised, studies whose effect sizes cannot be interpreted because the comparison groups are doing something we cannot characterise (e.g. Falchikov & Goldfinch, 2000; Panadero et al., 2023). Strong claims about the effects of self-assessment or peer assessment on student learning depend on research that is robust enough to support them, and the field does not yet produce enough such research.

The second problem concerns measurement. The field uses a range of measurement approaches that are not uniform across studies and not always justified within studies. What counts as a self-assessment measure, what counts as a measure of perception of self-assessment, what counts as a measure of accuracy or calibration, all of these vary considerably across the literature (Topping, 1998), and the variation is rarely discussed in ways that would allow synthesis across studies. The result is that meta-analyses in the field have to absorb a great deal of measurement heterogeneity, and the resulting effect sizes are noisier than they need to be. A more disciplined approach to measurement, with explicit articulation of what a given instrument measures and how it relates to instruments used elsewhere, would substantially improve the field's capacity to accumulate knowledge. The point is not that everyone should use the same instruments. It is that the limited use of well-validated measures (Brown & Zhao, 2023), coupled with insufficient psychometric reporting in published studies (often the instrument itself is not reproduced), makes cross-study comparison harder than in adjacent areas of educational research

where shared instruments exist. The question worth asking is what the field needs to converge on, even partially, in order to compare results across studies more meaningfully.

The third problem concerns the specification of interventions. When a study reports that students participated in a self-assessment intervention, what they actually did is often unclear from the published report. Did they apply criteria provided by the teacher, or did they develop criteria themselves? Did they receive feedback on their self-assessments, or were the assessments private? Was the practice integrated with grading or kept separate? Did students receive training, and if so, of what kind and for how long? Without this kind of specification (van den Berg et al., 2006), the comparison of one self-assessment study with another is difficult, and the claim that self-assessment has effects of a given size becomes a claim about a heterogeneous mix of practices that may have very different effects from each other. Better specification of what is actually being studied is one of the most important contributions individual researchers can make to the cumulative development of the field. The diagnosis is most useful when it is paired with concrete tools. One promising direction is the development of shared coding instruments that capture the components of self-assessment, peer assessment, and rubric interventions in enough detail to allow meaningful comparison across studies (osf.io/3hgcv; osf.io/5k42z; osf.io/u83eh). Instruments of this kind are the kind of contribution the field needs to address a problem that diagnosis alone cannot solve.

The fourth problem concerns how the field studies students' perceptions and experiences of assessment practice. There is a substantial literature on how students perceive self-assessment, peer assessment, feedback, rubrics, and related practices, and the literature has produced findings that are influential in informing pedagogical recommendations. But a careful look at this literature reveals that perceptions are studied in ways that are problematic in several respects. The measurement is rarely uniform across studies, the interpretation of perception data is sometimes more confident than the data warrants, and the connection between what students report perceiving and what they actually do in assessment activity is often left implicit (Brown & Zhao, 2023; Panadero et al., 2023). Perceptions matter in research on assessment, but they matter as one source of evidence among several, and the field has not always treated them with the methodological care they require. Working more carefully with what students actually do, alongside what they report perceiving, would strengthen the empirical base on which pedagogical claims rest.

These four problems are real, they cut across theoretical traditions, and they deserve sustained attention. Table 1 summarises each problem and what authors can report to address it. Reviewers and editors can use the same list when evaluating submissions. Several integrative models have been proposed in recent years that offer organisational frameworks for thinking about the components of self-assessment and peer assessment practice (e.g. Panadero et al., 2024; Yan & Brown, 2017; Yan & Carless, 2022). These models can be useful as ways of locating what a given study attends to and what it leaves out, and of clarifying which components require more careful empirical work. The detailed development of any one of these models is beyond the scope of the present paper. What I want to register here is that the field has the conceptual resources to organise its work, and the question is whether we use them to do the work or whether we use them to entrench positions. My hope is that the field uses them to do the work.

Table 1. Methodological problems in research on student involvement in assessment and recommendations for authors.

Problem	How it appears in the literature	What authors can do	Key references
Weak research designs	Studies without control conditions; comparison groups whose activities are not described; effect sizes that cannot be interpreted.	Include a comparison condition and describe what it involves. Operationalise the dependent variables. State which design features limit causal claims.	Falchikov and Goldfinch (2000); Panadero et al. (2023)
Non-uniform measurement	Self-assessment, perception, and accuracy measures vary across studies; psychometric evidence is rarely reported; instruments are often not reproduced.	Name the construct each instrument measures. Report validity and reliability evidence. Reproduce the instrument or deposit it in an open repository. Use established measures where they exist.	Brown and Zhao (2023); Topping (1998)
Underspecified interventions	Reports do not state what students actually did.	Report who set the criteria, whether students received feedback on their assessments, whether the practice counted toward grades, and the type and length of training. Code intervention components with a shared instrument.	van den Berg et al. (2006)
Uncritical use of perception data	Perception measures vary; interpretations go beyond the data; the link between what students report and what they do is left implicit.	Treat perceptions as one source of evidence. Pair self-reports with data on what students do, such as the assessments they produce. Do not infer learning effects from perceptions alone.	Brown and Zhao (2023); Panadero et al. (2023)

Doing the work includes testing these models. When synthesising research on assessment in higher education, I have often met the same pattern. A new framework is proposed and cited widely, and the next one replaces it before its claims have been tested. Mischel (2008) called this habit in psychology the *toothbrush problem*: the reluctance to use any theory but one's own. Frameworks then accumulate faster than the evidence that would test them. A framework should be judged by the studies it makes possible and by how its claims fare in those studies. This applies to the integrative models cited above as much as to any other. Panadero and Brown (2026b) develop the related point that reconceptualisations and empirical programmes answer to different standards.

Conclusion

The position I have articulated in this paper is not new. It is, I think, the position that some in the field have held in practice for some time, even when its public discourse has organised itself around polarities that are less productive than they look. What I have tried to do is name the position explicitly, articulate why I find it more useful than the polarised alternatives, and identify some concrete directions in which integrative work would be most valuable. The paper has been short because the position is itself relatively simple. The polarities are gradations. Positions in their debates can be judged by the integration test. The foundational work is owed engagement. The methodological problems are real and shared. The work of the field is to address them together. One concrete way to do so is to turn the recommendations in Table 1 into a consensus reporting guideline for research on student involvement in assessment. Such a guideline should be developed with the research communities that produce this work.

This position has been built over more than a decade of conversations with colleagues across several research communities, in particular EARLI SIG-1, AERA, the International Network for Formative Assessment, and I expect to continue developing it in the years ahead. The paper is offered in that spirit: as one contribution to a continuing conversation in a field that has more in common across its supposed divides than its narration sometimes suggests. The work is collective, the foundation is shared, and the most useful research the field can do is the research that takes both of those facts seriously.

Acknowledgments

An earlier version of the position developed in this paper was presented as a keynote lecture at the EARLI SIG-1 Assessment and Evaluation conference (Barcelona, June 27, 2024). The author thanks the SIG-1 community for the conversations over more than a decade in which these ideas have been built and tested. The author also thanks Gavin Brown and Anders Jönsson for comments on an earlier version of this paper.

Author contributions

CRediT: **Ernesto Panadero:** Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author.

The author used Claude Opus 4.8, 5, and 5.5 (Anthropic), a large language model, in preparing this manuscript. The tool was used to edit the text for clarity and concision and to check reference formatting and bibliographic details. The argument and all decisions about content are the author's. The author reviewed and revised all text produced with the tool, checked the references against the original sources, and takes full responsibility for the content of the article.

Funding

This work was supported by the Spanish Ministry of Science, Innovation and Universities under Grant [PID2024-156609NB-I00].

Notes on contributor

Ernesto Panadero is Director of the Centre for Assessment Research, Policy and Practice in Education (CARPE) at Dublin City University. His research focuses on self-assessment, peer assessment, feedback, rubrics, and self-regulated learning, much of it through systematic reviews and meta-analyses.

ORCID

Ernesto Panadero http://orcid.org/0000-0003-0859-3616

References

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, *4*, Article 87. https://doi.org/10.3389/feduc.2019.00087

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, *48*(1), 12–19. https://doi.org/10.1080/00405840802577544

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, *18*(4), 451–469. https://doi.org/10.1080/0969594X.2011.557020

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, *18*(5), 529–549. https://doi.org/10.1007/BF00138746

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 367–393). Sage.

Brown, G. T. L., & Zhao, A. (2023). In defence of psychometric measurement: A systematic review of contemporary self-report feedback inventories. *Educational Psychologist*, *58*(3), 178–192. https://doi.org/10.1080/00461520.2023.2208670

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*(4), 395–430. https://doi.org/10.3102/00346543059004395

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287–322. https://doi.org/10.3102/00346543070003287

Mischel, W. (2008, December). The toothbrush problem. *APS Observer*, *21*(11). https://www.psychologicalscience.org/observer/the-toothbrush-problem

Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment and Evaluation in Higher Education*, *48*(8), 1053–1075. https://doi.org/10.1080/02602938.2023.2164884

Panadero, E., & Brown, G. T. L. (2026a). Agency in self-assessment is a developmental achievement, not a starting point: Why prior knowledge and expert structure are preconditions for student evaluative capacity. *Assessment and Evaluation in Higher Education*. Advance online publication. 1–13. https://doi.org/10.1080/02602938.2026.2706014

Panadero, E., & Brown, G. T. L. (2026b). Recovering the foundational programme of student self-assessment: A response to Nieminen, Yan, and boud (2025). Assessment and Evaluation in Higher Education. Advance online publication. https://doi.org/10.1080/02602938.2026.2731625

Panadero, E., Fernández, J., Pinedo, L., Sánchez, I., & García-Pérez, D. (2024). A self-feedback model (SEFEMO): Secondary and higher education students' self-assessment profiles. *Assessment in Education: Principles, Policy & Practice*, *31*(3–4), 221–253. https://doi.org/10.1080/0969594X.2024.2367027

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education*, *30*(2), 175–194. https://doi.org/10.1080/0260293042000264262

Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, *34*(7), 807–826. https://doi.org/10.1080/03075070802706553

Sadler, D. R. (2013). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 5–19. https://doi.org/10.1080/0969594X.2012.714742

Sluijsmans, D. M. A., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment and Evaluation in Higher Education*, *27*(5), 443–454. https://doi.org/10.1080/0260293022000009311

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education*, *31*(3), 341–356. https://doi.org/10.1080/03075070600680836

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment and Evaluation in Higher Education*, *42*(8), 1247–1262. https://doi.org/10.1080/02602938.2016.1260091

Yan, Z., & Carless, D. (2022). Self-assessment is about more than self: The enabling role of feedback literacy. *Assessment and Evaluation in Higher Education*, *47*(7), 1116–1128. https://doi.org/10.1080/02602938.2021.2001431

Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, *37*, Article 100484. https://doi.org/10.1016/j.edurev.2022.100484