Commentary: Self-Report is Indispensable to Assess Students' Learning

Reinhard Pekrun[a]

[a]University of Essex, United Kingdom & Australian Catholic University, Sydney, Australia

Abstract

Self-report is required to assess mental states in nuanced ways. By implication, self-report is indispensable to capture the psychological processes driving human learning, such as learners' emotions, motivation, strategy use, and metacognition. As shown in the contributions to this special issue, self-report related to learning shows convergent and predictive validity, and there are ways to further strengthen its power. However, self-report is limited to assess conscious contents, lacks temporal resolution, and is subject to response sets and memory biases. As such, it needs to be complemented by alternative measures. Future research on self-report should consider not only closed-response quantitative measures but also alternative self-report methodologies, make use of within-person analysis, and investigate the impact of respondents' emotions on processes and outcomes of self-report assessments.

K*eywords*: self-report; emotion; motivation; metacognition; self-regulated learning
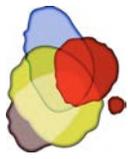
Introduction

Self-report is indispensable for any more nuanced assessment of mental states. While it is possible to examine general physiological properties of thought and affect using brain-imaging, and their consequences through performance tests and behavioral observation, assessing the contents and complex cognitive processes involved in human thinking, emotion, and motivation requires self-report. As such, self-report was a primary assessment method in psychology and education from early on, and it continued to be a primary method throughout all developmental phases in the history of these disciplines, even in the prime time of behaviourism early in the 20th century. However, self-report also has limitations. Self-report is restricted to processes that are accessible to consciousness; is typically limited to assess contents that can be verbally described; can be subject to various biases; and is always lagging behind the processes it aims to assess, even if only for seconds, which implies that it lacks the temporal resolution needed to capture the real-time dynamics of learning.

Given these problems, it is important to critically scrutinize the power of self-report methods to capture the constructs they intend to assess, and to develop strategies to improve their validity. The papers in this special issue are excellent examples for both directions. Specifically, all eight papers examine the validity of specific self-report instruments relative to proposed distributions of scores and relations with other variables. In addition, two of the papers also explore ways to improve validity. In the following sections, I first address the nature of self-report and its advantages and drawbacks. Next, I discuss the advances in analyzing and improving the validity of self-report measures that are represented in the contributions to this special issue. In conclusion, I outline three directions for future research.

1. What is self-report?

Self-report uses participants' verbal responses to assess their cognition, emotion, motivation, behavior, or physical state. When thinking about self-report, what often comes to mind first is structured questionnaires measuring some kind of personality trait. However, while structured questionnaires are used frequently, the most commonly employed self-report instrument likely is the clinical interview, which typically has a very different format as compared with closed-response questionnaires. By implication, to judge self-report, it is important to consider that this method can take very different forms. Self-report can be structured or unstructured; retrospective or concurrent; oral or written; qualitative of quantitative; one-dimensional or multi-dimensional; paper-and-pencil or online; and can comprise single or multiple items (see Pekrun & Bühner, 2014, for an overview). As such, self-report not only includes structured multi-item questionnaire scales, but also open-ended interviews, single-item momentary reports, unstructured think-aloud protocols, etc. While all these methods share properties of relying on participants' ability to self-assess and report about the variables under investigation, they differ vastly in terms of structure, temporal resolution, and metric used. As such, it is important to keep in mind that any findings on the validity of self-report instruments, and on ways to improve it, may be specific to some variant of self-report and not be generalizable to other variants.

In the current special issue, all of the eight contributions consider multi-item questionnaire scales using closed formats (i.e., defined items and response options). Rogier et al. (2020) additionally included a think-aloud protocol. As such, with this exception, the contributions focus on quantitative, structured self-report measures. Such measures are well suited to answer quantitative research questions that are defined a priori. However, they are less suited to answer exploratory questions and to gain a more nuanced picture of respondents' subjective world of multi-layered thoughts and perceptions, which can transcend researchers' prior conceptions as represented in closed-format scales. For such purposes, qualitative self-report methods are needed. Overall, to make progress in research on learning and instruction, it is often useful to employ a mix of quantitative and qualitative self-report, with qualitative methods used to explore new territory and gauge in-depth explanations, and quantitative method to test theoretical hypotheses in more rigorous ways (see, e.g., Pekrun et al., 2002).
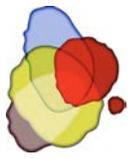
2. Benefits and drawbacks of self-report

Self-report has clear advantages. First, in contrast to other types of assessment, self-report allows assessment of all types of psychological processes. Observation can assess visible behavior, achievement tests can measure cognitive performance, neuro-imaging the activation of brain areas, and physiological measures the arousal of peripheral systems, but self-report can be used to assess all of the affective, cognitive, physiological, and behavioral processes that are part of self-regulated learning – all of these processes can be represented in the human mind and can be reported accordingly. Second, self-report can render a more differentiated assessment of human thinking than any other method. As such, for a nuanced description of emotions, motivation, and metacognition during learning, self-report is needed. Third, self-report is more economical than other methods. Self-report may be the only method applicable in some types of studies, such as large-scale student assessments.

Self-report also has disadvantages. As noted, self-report is limited to the assessment of processes that are accessible to consciousness. Responses that cannot be represented mentally need to be assessed with other methods. Another important limitation is the use of language (although self-report can also employ non-verbal communication). Research has shown that terms describing psychological processes tend to be used in consistent ways across languages (e.g., Fontaine et al., 2013), but there can nevertheless be differences in semantic understanding across cultures and learners. By implication, measurement equivalence of self-report instruments across groups should not merely be assumed but needs to be established empirically. Furthermore, limitations result from the fact that self-report is under respondents' control. Whereas it may be difficult to alter one's level of physiological activation, reports about perceived activation can easily be changed. As such, depending on motivation and preferences for response options, self-report can be subject to various response biases, such as social desirability.

Finally, self-report is also subject to memory biases. This is especially true for retrospective self-report that is administered at a later point in time and requires recollection of information from autobiographical memory, but is also true for state self-report asking respondents how they feel or what they think right now – self-report is lagging behind the phenomena it captures, even if only for seconds. As such, self-report inevitably lacks the temporal resolution needed to capture the real-time dynamics of psychological processes. This is even true for momentary methods such as experience sampling or think-aloud protocols as used in the contributions by Moeller et al. (2020) and Rogier et al. (2020). Even these methods cannot reach the temporal granularity of concurrent physiological or behavioral-observational methodologies. As such, self-report needs to be complemented or substituted with other methods for many research purposes. For making progress in research on learning, multi-channel assessments of motivation, emotion, and metacognition including self-report along with observational and physiological methods as well as behavioral trace data are especially promising (Azevedo et al., 2018; Lajoie et al., in press).

3. Examining the validity of self-report measures

Six of the eight contributions to this special issue focus on examining the validity of (quantitative) self-report measures and developing methods to examine validity. Iaconelli and Wolters (2020) investigated the impact of insufficient effort in responding on university students' self-report scores for their beliefs and behaviors during self-regulated learning. Self-report was assessed as part of students' coursework. Rates of insufficient effort were low, and reported relations between variables were robust against including students with insufficient effort, suggesting that inattentive responding does not represent a major threat to validity (at least under the situational conditions of the study).

Rogiers et al. (2020) examined secondary school students' retrospective self-report of learning
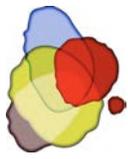
strategies used while learning from a text in combination with think-aloud data obtained during the same session. The data from the retrospective self-report were used to classify different types of learners, and the findings show that these types differed systematically in their learning process as assessed through the think-aloud protocol, thus attesting to the convergent validity of these two – very different – types of self-report instruments.

Extending the perspective beyond individual learning, Vriesema and McCaslin (2020) used self-report to assess secondary school students' general test anxiety, their attitudes towards school, and their behavior and emotions during group work in mathematics. There were clear links between self-reported behaviour and emotions related to the group work situation, but less so with the test anxiety measure. These findings are consistent with the specificity matching principle (see, e.g., Swann et al., 2007): Variables show stronger relations when being matched in terms of situational specificity than when not being matched, and the present results suggest that self-report measures can demonstrate validity when attending to this principle.

Van Halem et al. (2020) used the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991) as well as online trace data to assess undergraduate students' motivation and learning strategies during a statistics course. There was no direct conceptual match between the MSLQ and online trace data constructs. Nevertheless, there were substantial relations between time investment as assessed by the MSLQ, on the one hand, and trace data, on the other, thus demonstrating convergent validity of the two types of measures. Furthermore, both the MSLQ scores and the online trace data contributed to explaining students' course performance, thus supporting predictive validity for both types of measures.

Moeller et al. (2020) used experience sampling methodology (ESM) with a two-item measure of situational interest to capture the developmental dynamics of university students' interest in a series of lectures over one semester. In contrast to traditional ESM designs, a fixed rather than random schedule of assessments was used, which facilitated aggregation of assessments across participants. The findings of cross-classified multilevel analysis show that there was substantial variation of interest scores between students as well as within and between lectures, thus documenting the usefulness of situational self-report scores to decompose these sources of variance.

Finally, in terms of developing additional methods for testing validity, Chauliac et al. (2020) observed university students' gaze behavior while answering items on a questionnaire assessing habitual use of different cognitive strategies during learning from texts. There were systematic links between number and duration of fixations, on the one hand, and the consistency of answering different items from the same scale, on the other. The findings demonstrate that eyetracking has great potential to examine processes of responding to verbal stimuli as presented in self-report scales, suggesting that this methodology could contribute to examining the validity of scores derived from these scales.

Taken together, these six contributions attest to the potential validity of self-report in assessing students' learning. There were clear links (1) between quantitative self-report scores for different constructs, as well as (2) between these scores, on the one hand, and think-aloud protocols, online trace data, and academic performance, on the other. While not all of these links were fully robust and significant, they nevertheless document that self-report continues to be useful in measuring facets of students' learning.

4. Improving validity

How can we further improve the validity of self-report measures? Two of the contributions address this question. Fryer and Nakao (2020) examined the impact of type of response scale on levels, reliability, and factorial validity of self-reported task interest and its links with prior and subsequent domain interest in a sample of PhD students. Their study included two traditional formats (labelled categorical scale and visual analogue scale) as well as two more recent formats (slider and swipe scales).
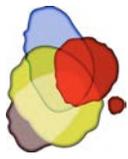
Reliability and factorial validity were nearly identical across the formats, and mean scores for interest in different tasks did not show systematic differences either. However, predictive validity for future interest tended to be higher for the slider and swipe versions than for the two traditional formats. This is promising and should stimulate research on how to further optimize response scales and their presentation.

Durik and Jenkins (2020) analyzed the link between undergraduate students' self-reported interest, their certainty in their answers, and their self-reported behavioral engagement in various subjects. The findings show that interest and certainty were related in a curvilinear fashion in most of the subjects; high certainty was associated with either low or high interest scores. Furthermore, the link between interest and behavioral engagement was substantially stronger for students with high certainty in their reported level of either individual or situational interest, and not even significant for students with low certainty in their situational interest. These findings suggest that including certainty ratings can increase the validity of self-report in predicting students' behaviour. As such, although replication is needed, they represent a potential breakthrough in boosting the validity of interest measures.

In both of these contributions, it remains open to question how the observed effects can be explained. For the effects of certainty, as noted by Durik and Jenkins (2020), it seems possible that clear beliefs in the strength or weakness of one's interest contributes to using interest as a guide for action, in contrast to being unclear about one's interest which may subject action to situational conditions. Research on the origins and outcomes of certainty is needed to examine this possibility.

5. Directions for future research

5.1 Scoping a broad range of self-report methods

The contributions to this special issue focus on self-report methods using written verbal statements as stimuli and closed-format options to respond, with quantitative methods employed to analyze responses. However, as noted earlier, there are various alternative formats that are equally important. Each of these formats has its own advantages and disadvantages. Specifically, a substantial amount of social science research uses oral formats (specifically, interviews) and open-ended answers, either written or provided orally. To an extent, these formats may be subject to similar biases as written closed-format self-report, including response sets and memory biases. However, there also may be differences, especially in terms of strategies to reduce these problems.

For example, motivation to respond in socially desirable ways rather than telling the truth can be reduced by generating trust in interviewees that their data will be kept confidential, and memory biases can be reduced through cognitive interviewing techniques that optimize recall. Substantial progress in suitable methods has been made in forensic psychology and research on testimony (see, e.g., Bowles & Sharman, 2014; Brown & Lamb, 2015). It would be worth exploring if some of these strategies could be made fruitful for educational research as well. This may be especially important for research on learning in young children (preschool, kindergarten, and the early elementary school years).

Qualitative self-report methodology using open-ended answers is especially important when exploring new research questions, but also when wanting to understand unexpected or paradoxical findings that can be explored with in-depth interviews. How to best structure questions, analyze answers, and aggregate qualitative self-report findings across studies currently is a field of intense methodological debate (see, e.g., Clark, 2016; Snelson, 2016). Mainstream quantitative research on self-report should attend to these developments, and research is needed on how to better integrate different self-report methods and the resulting evidence (e.g., in terms of convergent parallel, exploratory sequential, or explanatory sequential mixed-method study designs; Creswell, 2014; Creswell & Plano Clark, 2011).
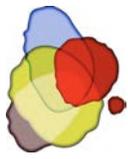
5.2 Importance of within-person research for validating self-report

Similar to other types of research in education and psychology, the vast majority of investigations using self-report have relied on between-person study designs, including most of the contributions to this special issue (the Moeller et al., 2020, contribution is a notable exception). Between-person research is suited to examine individual differences and interindividual relations between variables. However, it is not suited to investigate the within-person psychological functioning that is addressed in theories of students' motivation, emotion, and self-regulated learning. Intraindividual and interindividual correlations are statistically independent, and there is no easy way to infer one from the other, except when conditions of ergodicity hold. These conditions include homogeneity of functional relations across persons and stationarity over time, conditions that are often not met (Voelkle et al., 2014). As such, to study motivation, emotion, and strategy use during learning, it is best to examine these processes within persons (Murayama et al., 2017). To ensure generalizability, the variation of within-person relations across persons needs to be analyzed – if there is little variation, then relations are generalizable and nomothetic conclusions can be reached.

The relevance of within-person research has important implications for the validation of self-report measures. In research on learning, some of these measures pertain to trait-like characteristics of students and are used to gauge between-person differences. For example, measures of trait-like individual interest may be used to assess differences in interest between students. For these measures, it is appropriate to use between-person designs to examine validity. However, whenever the purpose is to assess individual development over time, or personal functioning during learning, then it is more adequate to use within-person designs to validate self-report methods. Between-person designs can render misleading conclusions, and resulting findings can under- or overestimate validity relative to theories of individual learning.

5.3 The role of emotion and emotion regulation in self-report

As human performance more generally, adequately responding to self-report instruments requires both competence and motivation. Competence includes being able to understand questions, to retrieve relevant information from long-term memory or current working memory, and to integrate the retrieved information such that a decision about an adequate answer can be reached. Current models of self-report largely focus on these cognitive processes, and process-oriented methods to validate self-report items focus on techniques of cognitive validation (Castillo-Diaz & Padilla, 2013; Karabenick et al., 2007). Motivation includes wishes to veridically answer questions, either to get a valid self-assessment (e.g., in contexts such as career counselling or psychotherapy) or to help researchers in their attempts to understand reality, as well as desires to appear to others or oneself as a socially desirable person. Motivation has been examined especially in research on social desirability (see, e.g., Gignac, 2013), and measures controlling for desirability have a long-standing tradition of being included in studies of personality.

However, beyond cognition and motivation, it seems likely that emotions also play a critically important role in self-report. Emotions are defined as affective responses to personally important events. As such, whenever self-report touches issues that are personally important, emotions are likely to be aroused. This can be emotions that are already associated with a given topic in memory, such as anxiety when retrieving recollections of prior exams, or emotions that are generated during the process of reading self-report items. In addition, emotions that are elicited by the task of responding and the social context of the assessment can play a role, such as sympathy for the experimenter administering a questionnaire, social anxiety of disclosing personal information, or anger about redundancy of items in lengthy multi-item instruments.

It is reasonable to assume that these emotions can substantially influence self-report responses. This can happen through the influence of emotions on retrieval of information from memory (e.g., in terms of mood-congruent retrieval), on integrating memory information in different ways (e.g., holistically in positive mood and detail-oriented in negative mood), on current motivation to persist in answering questions, and on motivation to answer in specific ways (e.g., according to social desirability when being socially anxious about one's responses). Furthermore, ways to regulate emotions may play
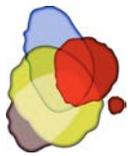
a role as well. For example, unpleasant emotions triggered by emotionally negative self-report items may be so strong and aversive that one seeks to downregulate them right away, even before answering the item. As a result, the answer may no longer represent the original emotional response to the item. Emotions can contribute to changes of the objects of self-report measurement during the process of measuring them – a phenomenon that can render resulting scores an artefact of the response process.

Research exploring these possibilities is largely lacking. Self-report methodologists could team up with memory researchers, social psychologist, and affective scientists to investigate these possible influences of emotions on self-report. The results could inform psychological and educational measurement in terms of shaping instruments and the social situations of assessment in ways that are both emotionally beneficial and suited to increase validity.

6. Conclusion

Self-report is indispensable for any more fine-grained assessment of mental processes, including students' motivation, emotions, cognitive strategies, and metacognition during learning. Certainly, self-report has limitations in terms of assessing conscious processes only, being subject to biases, and not providing the temporal resolution needed to assess some of these processes. Nevertheless, the evidence reported in the contributions to this special issue clearly document that self-report continues to be a valid way to assess processes of learning. To further boost its validity, triangulation of different self-report methods (such as closed items and open-format think-aloud protocols) as well as integration of self-report into multi-channel assessments can be helpful. To make further progress in examining and improving the psychometric quality of self-report methods, it may be useful to consider a broad range of different variants of self-report, to consider the influence of respondents' emotions on their self-report, and to complement traditional between-person study designs with intraindividual analysis.

Keypoints

- Self-report is required for a nuanced assessment of mental processes
- Self-report is indispensable to assess learning-related emotions, motivation, meta-cognition, and self-regulation
- Learning-related self-report scales show convergent and predictive validity
- Self-report needs to be amended by alternative measures because it lacks temporal resolution and is subject to response sets and memory biases
- Future research should consider a broader range of self-report methods, within-person analysis, and the impact of emotions on self-report

References

Azevedo, R., Taub, M., & Mudrick, N.V. (2018). Using multi-channel trace data to infer and foster self-regulated learning between humans and advanced learning technologies. In D. Schunk & Greene, J.A (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 254-270). Routledge.

Bowles, P. V., & Sharman, S. J. (2014). A review of the impact of different types of leading interview questions on child and adult witnesses with intellectual disabilities. *Psychiatry, Psychology and Law, 21*, 205–217. https://doi.org/10.1080/13218719.2013.803276
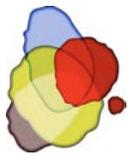
Brown, D. A., & Lamb, M. E. (2015). Can children be useful witnesses? It depends on how they are questioned. *Child Development Perspectives, 9*, 250-255. https://doi.org/10.1111/cdep.12142

Castillo-Diaz, M., Padilla, J.-L. (2013). How cognitive interviewing can provide validity evidence of the response processes to scale items. *Social Indicators Research, 114*, 963–975. https://doi.org/10.1007%2Fs11205-012-0184-8

Chauliac, M, Catrysse, L. , Gijbels, D. & Donche V. (2020). It is all in the surv-eye: can eye tracking data shed light on the internal consistency in self-report questionnaires on cognitive processing strategies? *Frontline Learning Research*, *8*(3), 26 – 39. https://doi.org/10.14786/flr.v8i3.489

Clark, A. M. (2016). Why qualitative research needs more and better systematic review. *International Journal of Qualitative Methods, 15*, 1-3. https://doi.org/10.1177/1609406916672741

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage.

Durik, A. M. & Jenkins J. S. (2020). Variability in Certainty of Self-Reported Interest: Implications for Theory and Research. *Frontline Learning Research*, *8*(3) 85-103. https://doi.org/10.14786/flr.v8i3.491

Fontaine, J. J. R., Scherer, K. R., & Soriano, C. (Eds.). (2013). *Components of emotional meaning: A sourcebook*. Oxford University Press.

Fryer, L. K., & Nakao K. (2020). The Future of Survey Self-report: An experiment contrasting Likert, VAS, Slide, and Swipe touch interfaces. *Frontline Learning Research, 8*(3), 10-25. https://doi.org/10.14786/flr.v8i3.501

Gignac, G. E. (2013). Modeling the Balanced Inventory of Desirable Responding: Evidence in favor of a revised model of socially desirable responding. *Journal of Personality Assessment, 95*, 645–656. https://doi.org/10.1080/00223891.2013.816717

Iaconelli, R., & Wolters C.A. (2020). Insufficient Effort Responding in Surveys Assessing Self-Regulated Learning: Nuisance or Fatal Flaw? *Frontline Learning Research*, *8*(3) *104 – 125.* https://doi.org/10.14786/flr.v8i3.521

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Ree Bonney, C.,…& Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*, 139–151. https://doi.org/10.1080/00461520701416231

Lajoie, S. P., Pekrun, R., Azevedo, R., & Leighton, J. P. (in press). Understanding and measuring emotions in technology-rich learning environments. *Learning and Instruction*.

Moeller, J., Viljaranta, J., Kracke, B., & Dietrich, J. (2020). Disentangling objective characteristics of learning situations from subjective perceptions thereof, using an experience sampling method design. *Frontline Learning Research*, *8*(3), 63-84. http://doi.org/10.14786/flr.v8i3.529

Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. In D. W. Putwain & K. Smart (Eds.), *British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education – Current Trends: The Role of Competence Beliefs in Teaching and Learning* (pp. 71-87). Wiley.

Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 561-579). Taylor & Francis.
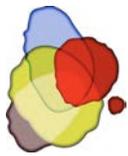
Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*, 91-106. https://doi.org/10.1207/S15326985EP3702_4

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)* (Tech. Report No. 91-B-004). Board of Regents, University of Michigan, Ann Arbor, MI.

Rogiers, A., Merchie, E., & Van Keer, H. (2020). Opening the black box of students' text-learning processes: A process mining perspective. *Frontline Learning Research*, *8*(3) 40 – 62. https://doi.org/10.14786/flr.v8i3.527

Swann Jr, W. B., Chang-Schneider, C., & McClarty, K. L. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist, 62*, 84–94. https://doi.org/10.1037/0003-066X.62.2.84

Van Halem, N., van Klaveren, C., Drachsler H., Schmitz, M., & Cornelisz, I. (2020). Tracking Patterns in Self-Regulated Learning Using Students' Self-Reports and Online Trace Data. *Frontline Learning Research*, *8*(3), 140-163. https://doi.org/10.14786/flr.v8i3.497

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Towards a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research, 49*, 193–213. https://doi.org/10.1080/00273171.2014.889593

Vriesema, C.C., & McCaslin, M. (2020) Experience and Meaning in Small-Group Contexts: Fusing Observational and Self-Report Data to Capture Self and Other Dynamics. *Frontline Learning Research*, *8*(3), 126-139. https://doi.org/10.14786/flr.v8i3.493