Practical Assessment, Research, and Evaluation

Volume 26 Article 7

2021

Do Loss Aversion and the Ownership Effect Bias Content Validation Procedures?

Vanessa Svihla
University of New Mexico

Amber Gallup
University of New Mexico

Do Loss Aversion and the Ownership Effect Bias Content Validation Procedures?

Cover Page Footnote

This material is based upon work supported by the National Science Foundation under Grant No. EEC 1751369. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We would also like to acknowledge the expert reviewers who put their time, effort, and expertise into evaluating our survey, and Jay Parkes, who provided the inspiration for pursuing this article.

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 26 Number 7, March 2021 ISSN 1531-7714

Do Loss Aversion and the Ownership Effect Bias Content Validation Procedures?[1]

Vanessa Svihla, *University of New Mexico*
Amber Gallup, *University of New Mexico*

In making validity arguments, a central consideration is whether the instrument fairly and adequately covers intended content, and this is often evaluated by experts. While common procedures exist for quantitatively assessing this, the effect of loss aversion—a cognitive bias that would predict a tendency to retain items—on these procedures has not been investigated. For more novel constructs, experts are typically drawn from adjacent domains. In such cases, a related cognitive bias, the ownership effect, would predict that experts would be more loss averse when considering items closer to their domains. This study investigated whether loss aversion and the ownership effect are a concern in standard content validity evaluation procedures. In addition to including promising items to measure a relatively novel construct, framing agency, we included distractor items linked to other areas of our evaluators' expertise. Experts evaluated all items following procedures outlined by Lawshe (1975). We found on average, experts were able to distinguish between the intended items and distractor items. Likewise, on average, experts were somewhat more likely to reject distractor items closer to their expertise. This suggests that loss aversion and the ownership effect are not likely to bias content validation procedures.

Introduction

Gathering evidence of validity is central to survey and assessment development, whether the constructs being measured are well understood or relatively novel. Although commonly described as such, an instrument is not "valid" or "validated"; instead, we evaluate whether the information acquired from an instrument is valid for particular uses (AERA, APA, & NCME, 2014; Kane, 2001; McCoach, Gable, & Madura, 2013; Sireci & Faulkner-Bond, 2014). One form of evidence for such arguments (Kane, 2001) is based on the instrument's content—whether the items adequately and fairly assess what the instrument is intended to measure (Lynn, 1986; Salkind, 2010). Traditionally termed *content validity*, (AERA et al., 2014), we follow the approach taken elsewhere of using this term, but anchoring it to evidence and purpose (e.g., Sireci & Faulkner-Bond, 2014).

The central challenge of evaluating the validity of content is fully capturing "the construct without

[1] This material is based upon work supported by the National Science Foundation under Grant No. EEC 1751369. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We would also like to acknowledge the expert reviewers who put their time, effort, and expertise into evaluating our survey, and Jay Parkes, who provided the inspiration for pursuing this article.

bringing extraneous constructs into the operationalization" (McCoach et al., 2013, p. 93). Although modern views of validity outline a range of procedures that may be used to gather evidence about the validity of content depending on the purpose of the instrument and how it will be used (Kane, 2016), a common technique involves expert review and judgment (AERA et al., 2014; McCoach et al., 2013; Sireci & Faulkner-Bond, 2014). Many such procedures are based in Lawshe's (1975) approach (Howard, 2018; Wilson, Pan, & Schumsky, 2012). Lawshe sought to establish a means to measure the overlap between actual job performance and various measures of job performance. He tasked individuals with independently assessing items as essential, useful but not essential, and not necessary to perform the particular job. Such assessments depend on the expertise of the individuals, as the evaluation of the item fit is based on whether their independent assessments demonstrate consensus. Thus, selecting qualified experts is critical. In the 2014 *Standards for educational and psychological* testing, Standard 1.9 details that requisite qualifications and/or experience should be reported, as there exists no uniform metric for selecting experts (AERA et al., 2014). A long-standing gap in procedures used to gather evidence about content validity is systemic evaluation of the quality of expert judgements (Stelly & Goldstein, 2007), as these are subjective (Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003). Others have raised concerns about similar procedures based on findings that experts' judgment can vary depending on the nature of the task, with more routine tasks likely to receive more consistent evaluation (Wyse & Babcock, 2018).

Insights from cognitive science also raise concerns about typical content validity evidence procedures—namely, individuals commonly make decisions to avoid loss (Tversky & Kahneman, 1991). If loss aversion affects experts as they evaluate survey items, they may be biased toward retaining items that should not be retained. Studies suggest experts and novices alike are affected by loss aversion (Kühberger, 1998; Loke & Tan, 1992). However, experts may also be impacted by the ownership effect. Individuals are more averse to parting with an object if it is similar to one they own (Morewedge, Shu, Gilbert, & Wilson, 2009). As such, evaluating items close to one's expertise could induce this effect. Afterall, experts can become territorial when someone they perceive as having less expertise questions their judgement about museum object social tags (Thom-Santelli, Cosley, & Gay, 2010) and Wikipedia entries (Thom-Santelli, Cosley, & Gay, 2009). Such territoriality, which clearly extends beyond the physical into psychosocial domains like expertise (Kirk, 2017), depends on sense of ownership (von der Trenck, 2015).

This ownership effect would predict that the more similar to their own research expertise the item is, the more risk averse they may behave; in other words, they may favor retaining items closer to their expertise, even if the items do not align well to the construct being measured. For well-established constructs, we would predict that experts would be relatively consistent in the degree to which they are impacted by sense of ownership, assuming they have similar expertise. However, for less established constructs, guidelines suggest recruiting experts from adjacent areas (Davis, 1992). In such cases, the ownership effect may be revealed, as each expert would be predicted to show loss aversion with items closest to their expertise.

The purpose of this study was to investigate whether loss aversion and sense of ownership explain expert behavior in evaluating potential survey items for a relatively novel construct, framing agency. Framing agency was recently characterized as a means to differentiate the kinds of decisions that matter in learning (Svihla, Gomez, Watkins, & Peele-Eady, 2019; Svihla & Peele-Eady, 2020). Learners make many decisions—whether to attend class, what to wear to class, how to engage during class, etc. However, most of these decisions have little bearing on *how* they learn, because the instructor typically directs this. Yet, in some learning settings, such as project-based learning, makerspaces, and design projects, students make decisions that are consequential to how they will learn about the problem because they make decisions about how the problem is framed—meaning they display framing agency. Our past research, conducted using discourse analysis, has highlighted that framing agency is detectable even early in design work and is instrumental to learning how to design (Svihla et al., 2019; Svihla & Peele-Eady, 2020). Students who display framing agency treat the problem as endemically constrained yet malleable, treat ideas as tentative, and make decisions that are consequential to the problem frame, yet share their agency with co-designers, materials, and stakeholders (Svihla, Gallup,

& Kang, 2020). In developing a survey to measure framing agency, we recognized that there were few true experts as the construct is rather new.

Historically, agency has been treated as context independent, similar to early accounts of constructs like self-efficacy, which is now commonly studied in context. For instance, self-efficacy has been studied as confidence in ability: to understand and solve engineering design problems (Avsec & Szewczyk-Zakrzewska, 2018), to be successful in academic settings (Pajares, 1996), to teach using inquiry methods (Richardson & Liang, 2008), and many others. When extending self-efficacy to a new context, it is relatively straightforward to select a panel of experts, which would include those with expertise in self-efficacy—but in another context—as well experts and practitioners with salient experience in the context. In contrast, agency has been investigated as an individual attribute, as shared with others, and as constrained by situations (Ahearn, 2001; Bratman, 2013; Narayan & Petesch, 2007), but not as contextualized by them. In contrast to the ways self-efficacy has been contextualized to relatively broad or domain-specific contexts like engineering design or teaching inquiry, we reasoned that decision making might be best contextualized by the nature and consequentiality of the decisions. Contextualizing agency in this way was novel. Thus, in seeking to recruit experts to assess whether items adequately and fairly measured framing agency, we recognized that true experts did not exist outside our research group.

In both the self-efficacy and framing agency examples above, experts from adjacent domains are involved. As such, loss aversion paired with sense of ownership over adjacent topics—areas in which our experts held deep expertise—could shape their decisions about items. Because extending self-efficacy to new contexts is a longstanding practice, an adjacent expert might interpret their role in a narrower manner, focusing their assessments on aspects of the items in their area. We argue that in the latter case of framing agency, where contextualization of agency is not common, we are afforded an ideal case to examine how evaluators' other areas of expertise shape their judgements. In this study, we sought to investigate the extent to which experts' evaluations of survey items display loss aversion and the ownership effect.

Most content validity evidence assessments include the following steps (Dillman, Smyth, & Christian, 2016), using either the full set or a subset of possible items: First, identify experts to review the items. The number of experts suggested when forming a panel varies, with several studies suggesting at least three experts (DeVellis, 2016; Gable & Wolf, 1993; Lynn, 1986; Rubio et al., 2003). Their qualifications may be based on academic or research as well as practical experience, depending on the intended use of the instrument (Grant & Davis, 1997). Second, orient experts to the constructs of study. In the case of more novel constructs, providing clear definitions is needed (Grant & Davis, 1997). Forming these definitions may itself be the subject of an expert review process; such reviews evaluate the adequacy of the definition that describes the content or construct to be measured, its sub-categories or subconstructs and their levels, and other guiding standards as applicable (Sireci & Faulkner-Bond, 2014).

Third, experts assess each item using a relevance scale—often a 3- or 4-point Likert scale ranging from not relevant to highly relevant (Davis, 1992). This may include questions about whether the items adequately cover the content or construct being measured (Sireci & Faulkner-Bond, 2014).

Fourth, the agreement between ratings is calculated for each item. Originally proposed as the content validity ratio (Lawshe, 1975), most methods evaluate the percentage of experts arguing to retain each item. For instance, the content validity index (CVI) has received considerable attention (Davis, 1992; Grant & Davis, 1997; Lynn, 1986; Martuza, 1977). CVI can be calculated for each item as the percentage of evaluators rating the item as a 3 or 4 on a 4-point scale. This approach collapses the 4-point scale into two values (Polit & Beck, 2006). Importantly, with more experts, the probability of complete agreement goes down (Rubio et al., 2003), meaning the cut score for retaining an item should be dependent on the number of experts reviewing it (Lynn, 1986). Fifth, assess the instrument CVI, often calculated as an average of item-level CVI (Polit & Beck, 2006), in part because the percent of items on which all experts agreed on retention decreases with a larger number—more than 5—of experts (Davis, 1992; Grant & Davis, 1997), an approach that is suggested to reduce the impact of a rater who is more or less critical on average (Haynes,

Richard, & Kubany, 1995; Lynn, 1986). Others have raised concerns that basing decisions on such small numbers of experts—a standard and common practice—reduces the precision (Beckstead, 2009), thus making the study of particular sources of potential systematic bias salient.

Method

Study design

The purpose of this study was to assess the degree to which loss aversion and sense of ownership might affect expert assessment of survey items. To evaluate this, we included distractor items on a survey as part of standard content validity evidence procedures. While this is not a common approach in content validation—in fact, we could find no other instances of this approach reported—we drew inspiration from the commonplace technique of including distractors on multiple-choice exams, which typically serves as a means to differentiate between accurate and inaccurate conceptions of particular content (Ali, Carr, & Ruit, 2016). Specifically, we posed the following research questions:

- Loss aversion: Do experts suggest retaining items, even those items not related to the intended construct?
- Ownership effect: Do experts preferentially suggest retaining items related to their other areas of expertise, but not related to the intended construct?

Participants

We identified a pool of ten experts in engineering design to review the Framing Agency Survey. We first identified six experts we personally knew through our research activities (e.g., reading and citing their work, interacting at conferences), and second, identified seven scholars who showed strong interest in framing agency when speaking to us at conferences. From this list, we carefully reviewed their publication records to evaluate whether they had relevant expertise and experience related to design or agency, eliminating three who did not. We reviewed the publication records of the remaining 10 experts to establish their additional areas of expertise, which included self-efficacy, identity/interest development, motivation, and self-regulated/directed learning.

Six experts returned the full survey; this included three we knew personally and three we had identified at conferences. The experts were all university professors of engineering or engineering education in the United States. These reviewers share several characteristics and experiences: all are designers themselves and have taught design at the university level; all are experienced in the development, adaptation, and use of surveys in research. Five of the six have conducted research on design.

Materials and procedures

We developed the Framing Agency Survey to measure undergraduate students' framing agency; that is, their consequential decision-making about how to frame design problems. The development process for the Framing Agency Survey is described in detail elsewhere (Svihla et al., 2020) following typical survey development strategies and validation efforts, including literature review and grounding items in data, organizing them by subconstructs (shared versus individual agency, ill-structuredness & tentativeness, constrainedness, and consequentiality of decisions, see sample items in Table 1), developing a large initial set of items, pilot testing items for word choice clarity through think aloud protocols with students, expert review to gather content validity evidence, and pilot testing with exploratory factor analysis (Dillman et al., 2016). This study re-evaluates the data collected through expert review; thus, we provide detail about these procedures.

We developed instructions, a definition of framing agency, and a scoring sheet. We introduced the task by explaining "We are developing the survey in order to learn more about how students develop and exercise framing agency in course-based design experiences and about the characteristics of learning experiences that support students' framing agency. We have asked you to assist us because your expertise is valuable as we evaluate and validate the survey content." Thus, we sought to link the task to their expertise and did not inform them of the deliberate inclusion of distractor items.

The definition was one full page and answered a series of questions: What is framing agency? Why do we need the construct of framing agency? What

differentiates students who show or do not show framing agency? The scoring sheet provided the full question text, a description of the response (e.g., "7-point scale ranging from very free to very limited"), four boxes for rating the relevance (1 = Not relevant; 2 = Somewhat relevant; 3 = Quite relevant; 4 = Very relevant), and a field for comments. At the end of the scoring sheet, we included a short expertise survey, asking them to provide a checkmark next to any that were true: I have developed a survey in the past, I have adapted a survey in the past. I have taught survey development. I have used a survey in research. I have no experience with surveys. I am a designer. I have taught design. I have studied design. I have done research on design. I have no experience with design.

We developed 52 items related to four subconstructs of framing agency as well as 17 distractor items (Table 1). Based on reviews of publications by the initial pool of 10 experts, we characterized four areas of partially overlapping expertise: identity and interest development; self-directed/self-regulated learning; motivation; and self-efficacy. None of the experts who completed the survey had published studies focused on self-efficacy or on three questions that were included but not grouped by construct and unrelated to any of our experts' areas. For each of these sets, we chose to generate three or four questions that were similar but not identical to existing items. We chose this number to enable some variability without over-burdening them. We wanted all experts to review all items to allow us to make comparisons between their judgements of items that were related to framing agency, to areas clearly in their expertise, and to areas outside their expertise.

Table 1. Sample items related to the framing agency subconstruct of individual consequentiality and distractor questions linked to reviewer expertise beyond design. Three additional questions were not grouped by construct.

Stem and questions	*Construct*
How responsible or not responsible did you feel for the outcomes of the design project? How responsible or not responsible did you feel for making decisions personally?	Framing agency: individual consequentiality
How interesting or uninteresting did you personally find this design problem? How interesting or uninteresting do you personally find this course as a whole? How interested or uninterested are you in pursuing engineering as a career?	Interest / Identity (4 experts, 3 items)
Considering your design project, what was easy or challenging for you? • Ordering the design tasks. • Locating information to solve the problem • Agreeing on the appropriate solution • Reporting on the solution	Self-direction / self-regulated learning (3 experts, 4 items)
How high or low would you rate your personal motivation to complete the design challenge? To what extent do you agree or disagree with the following statements? • I am excited about taking this course. • I am satisfied with how I am achieving my educational goals. • I start each school year highly motivated, and I maintain this motivation throughout the year.	Motivation (2 experts, 4 items)
I am confident I can: • Finish a class project as part of a team. • Communicate respectfully with classmates during a team class project, even when others disagree with me. • Amicably resolve disputes that may arise during a team class project.	Self-efficacy (0 experts, 3 items)

Via email, we provided the experts with a definition of framing agency and its sub-constructs, a link to the first author's website that provided additional project information, a copy of the Framing Agency Survey formatted as it would be for participants, a set of review instructions, the scoring sheet, and a publication in which we characterized framing agency based on design team discourse.

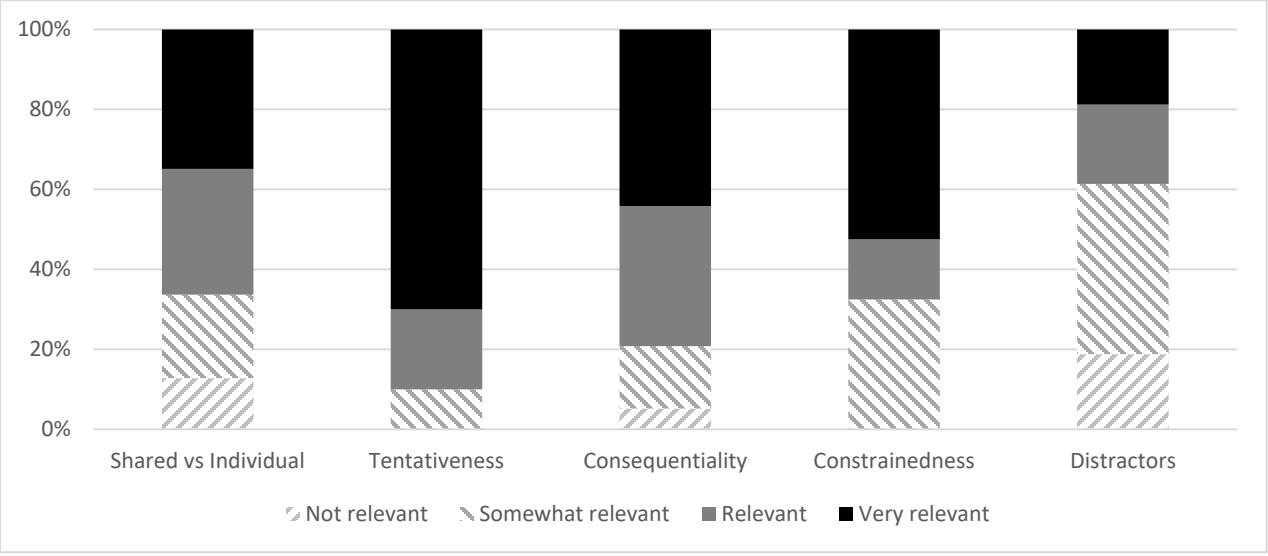
Data analysis

We calculated the percent agreement by question, by construct, and by grouping all distractor questions compared to the framing agency constructs. We calculated descriptive statistics for the four distractor constructs (motivation, identity/interest development, self-direction/regulation, self-efficacy). Using descriptive statistics, we compared the ratings of those with and without expertise in three distractor constructs (motivation, identity/interest development, self-direction/regulation). We reviewed comments related to these constructs. Finally, we compared the judgements made on distractors within and outside each evaluator's areas of expertise.

Results

Overall, the experts rated the distractor items as not or somewhat relevant, suggesting they were collectively able to differentiate between the items intended to measure framing agency and those introduced as distractors (Figure 1). They also differentiated between questions that fit framing agency subconstructs well and poorly. This suggests that the inclusion of distractors did not prevent them from using the full range of the scale when considering the intended items. We shared these results in more detail elsewhere (Svihla et al., 2020), but in brief, based on expert review, of the 52 framing agency items, we retained 25 Likert items for pilot testing, which resulted in clear latent variables tied to 18 items. We removed an additional seven cross-loaded items, all of which had received mixed reviews by experts. That experts suggested removing many items, and that remaining items generally loaded on distinct factors in pilot testing suggest the experts were not impacted by loss aversion overall. Had they been affected, they would have suggested retaining more items that, in pilot testing, might not have loaded on clear subconstructs.

Figure 1. Percent of expert judgements of the 52 framing agency and 17 distractor items as very relevant, relevant, somewhat relevant, and not relevant. Items are grouped by framing agency subconstructs, with all distractors grouped together.



Two experts had previously published multiple studies that included motivation as a major focus. Four distractor questions focused on motivation, receiving an average rating of 2.2 (*SD* = 0.9, a score suggesting they should be omitted) across judges, and an average score of 1.75 (*SD* = 0.5) from those with and an average of 2.4 (*SD* = 1.0) from those without expertise in motivation (Figure 2). One without expertise suggested retaining all four questions, citing that such questions would be useful in "post hoc analysis." One expert likewise suggested there could be "contextual value" to measuring motivation, but advised that better scales existed for this.

Four experts had previously published multiple studies that included identity and interest development. Three distractor questions focused on identity and interest, receiving an average score of 2.6 (*SD* = 1.1) across judges, and an average score of 2.6 (*SD* = 1.2) from those with and an average of 2.6 (*SD* = 0.6) without expertise in identity (Figure 2). Two with expertise suggested retaining all items, explaining that interest affects engagement. Two without expertise suggested retaining some items but did not elaborate on their reasoning. These scores, on the borderline between retention and omission for both groups, may reflect that such constructs seem salient when considering framing agency. Indeed, theory has long linked identity and agency (Holland, Lachicotte, Skinner, & Cain, 1998; Lave & Wenger, 1991).

Three experts had previously published multiple studies that included self-direction or self-regulation as a focus. Four distractor questions focused on self-direction/self-regulation, receiving an average score of 2.5 (*SD* = 0.2, a score suggesting they should be omitted) across judges, an average of 1.8 (*SD* = 0.8, a score suggesting they should be omitted) from those with and an average of 3.2 (*SD* = 0.9, a score suggesting they should be retained) from those without expertise in self-direction/self-regulation (Figure 2). The three without expertise suggested retaining most of these items, whereas only one with expertise recommended retaining just one item from this set. Those without expertise in this construct did not elaborate on their reasoning. Those with expertise in this construct explained that they could not see a connection between these questions and framing agency.

None of the experts had previously published more than one study that included self-efficacy as a focus. Three distractor questions focused on self-efficacy, receiving an average score of 2.2 (*SD* = 1.0, a score suggesting they should be omitted) across judges (Figure 2). Two suggested retaining all items, arguing that self-efficacy to resolve disputes and communicate respectfully could contribute to "collaborative sensemaking." Others recognized the questions as assessments of self-efficacy and therefore not direct measures of agency. The three remaining distractor questions that did not group by construct were evaluated similarly, receiving an average score of 2.5 (*SD*=1.0, a score suggesting they should be omitted).

Figure 2. Average evaluations by those with and without expertise (based on publication record) in distractor constructs of motivation, identity/interest development, self-regulation/direction, and self-efficacy. Error bars are standard deviation.

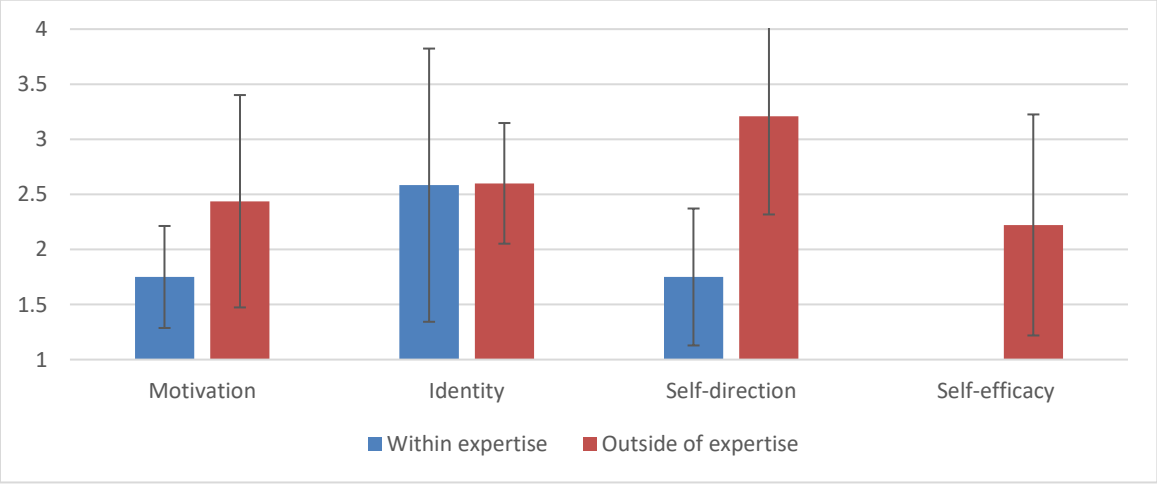
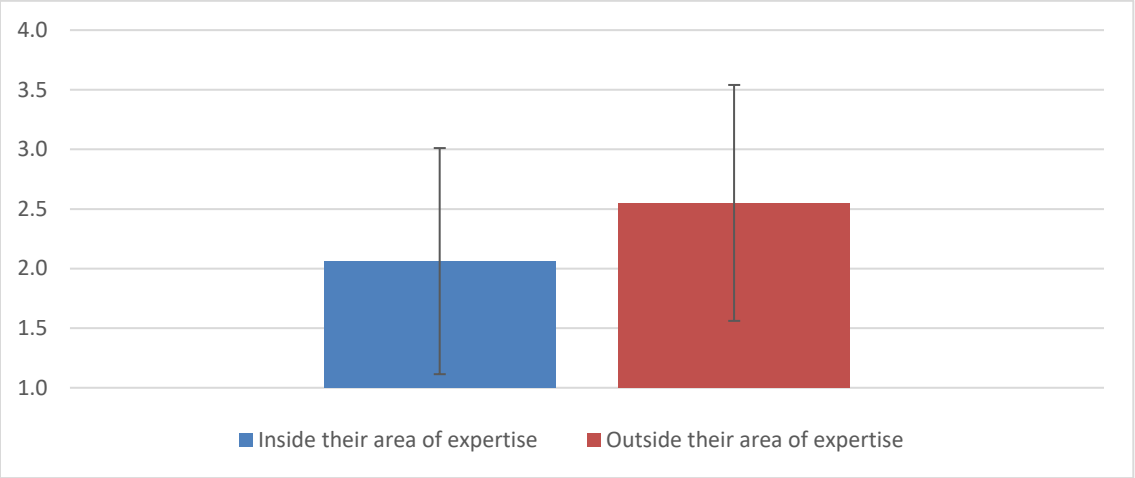


Figure 3. Average aggregate scores (across experts and distractor items) when evaluating items within their broader expertise and outside of their expertise. Error bars are standard deviations.



Overall, these results suggest that the ownership effect had little impact on experts' judgements. In the cases of motivation and self-direction/regulation, having expertise appears to have aided the experts in recognizing these items as different from framing agency. Likewise, those who recognized the self-efficacy questions as such used their knowledge to rule these questions out.

Looking across the total set of judgments of distractors, there were 35 instances where an evaluator made a judgement within their area of expertise and 66 outside of their area of expertise. We compared their scores, finding that those within their area of expertise (Figure 3, $M = 2.1$, $SD = 0.95$) were on average lower than those outside their area of expertise ($M = 2.6$, $SD = 1.0$). Thus, rather than showing an ownership effect, we found that experts tended to use their expertise to differentiate and suggest omitting items not related to framing agency.

Conclusions and discussion

We sought to explore whether loss aversion or the ownership effect biased reviewer judgements during typical content validity evidence procedures. Overall, we found that experts were able to clearly differentiate between items intended to measure framing agency and distractor items. This suggests that their judgements were not biased by loss aversion. Past research has suggested that the effects of loss aversion can bias decision making (Montibeller & Von Winterfeldt, 2015). In other settings, the risk of losing something typically outweighs the potential to gain something (Tversky & Kahneman, 1991). Perhaps in the case of evaluating prospective survey items, evaluators perceived little risk. By framing the task as dependent on their expertise, they may have felt empowered to suggest omitting questions. Because our experts also had experience with survey development, they may also have viewed survey fatigue as a greater risk. In other research methods that rely on expert judgement, a range of biases have been shown to affect decision-making (Bonaccorsi, Apreda, & Fantoni, 2020). For instance, in Delphi methods, experts' judgements are vulnerable to desirability bias (Ecken, Gnatzy, & Heiko, 2011). Thus, this study extends research on cognitive biases at play in studies reliant on subjective expert judgements.

In other settings, loss aversion appears to affect both novices and experts similarly (Kühberger, 1998; Loke & Tan, 1992). However, ownership effects can induce a stronger sense of loss aversion when evaluating objects similar to those one personally owns (Morewedge et al., 2009). We wondered if, when evaluating items closer to their other areas of expertise, evaluators might be subject to ownership effects. By comparing experts' judgements on items associated or not with other areas of their expertise, we found no evidence of this effect. On the contrary, experts used their judgment to suggest omitting items that were related to their expertise but not to the study

constructs. This finding is particularly promising for those developing instruments to measure relatively novel constructs, where it is not possible to form a panel of evaluators with expertise in the specific construct. However, we also found that one distractor construct—identity and interest development—was perceived by our evaluators as salient for understanding framing agency, perhaps reflecting theories that link agency and identity (Holland et al., 1998). Based on our results, we recommend including a secondary scale for experts to use to articulate that an item may be related or even predictive, yet not relevant to measuring the construct of interest.

One limitation of our methods was assessing each evaluator's expertise based on their publication record. Omitted from this record are areas in which they teach, industry experience, and projects in development. We selected this approach to avoid prompting the experts to look for distractor constructs in the items. Future research could incorporate a multi-phase study design to assess each expert's depth of knowledge related to each distractor construct.

Because this study was conducted in tandem with efforts to evaluate construct validity for a relatively novel construct, our sample size was small, as is typical with such methods. While this study provides initial evidence that loss aversion and the ownership effect were not an issue for our content validity evidence procedures, more work is needed to determine how systematic this is. We could argue that our findings extend to well-established constructs, where judges use their expertise to make nuanced decisions about item fit and coverage. However, we caution against this without further study, as the similarity of judgements across experts would potentially mask any bias present. Future studies could investigate this by asking a larger sample of experts to evaluate sets of questions that are in their field, adjacent to their field, and further from their field. Likewise, comparisons of performance in the presence and absence of distractor items could investigate the impact that including distractors has. It is possible that including distractors could make it more difficult for evaluators to differentiate between the best and worst fitting items that are more closely related to the construct. Alternatively, including distractors could make it easier for evaluators to overcome loss aversion effects, if they deem items to be clearly misfit.

Many practices in education, certification, and licensure depend on evidence of content validity gathered through procedures similar to those we used. For instance, when new curricular standards are adopted, authorizing agencies such as national, state, or district education departments may develop novel instruments to assess teacher implementation. Curriculum developers commonly develop instruments to evaluate fidelity as well as various impacts beyond learning. As we continue to expand our understanding of social-emotional supports and ways these explain variance in learning and assessments, new measures are needed to not only characterize learners, but also the degree to which educators are able to understand and implement effective programs. And as we continue to expand the ways we use technology in education, from communication with families to supporting learning, new instruments will be needed to shed light on the various impacts of these. Many such efforts share the characteristic of relative novelty noted for framing agency. Thus, understanding whether experts from adjacent domains are vulnerable to loss aversion and the ownership effect as they make judgements is critical for enhancing the trustworthiness of systems depending on the information new instruments might provide. Collectively, our results affirm the commonplace practice of using a small number (three to six) of adjacent experts in content validity evidence procedures for relatively novel constructs. Had the cognitive biases of loss aversion and the ownership effect been detected, we would have followed with additional field testing to identify ill-performing items or we might have conducted a training or calibration session with experts to reduce the ownership effect by ensuring they understood framing agency and how it differed from other areas of their expertise. Our results suggest these more burdensome strategies were not necessary in our case. However, keeping in mind that validity is tied to how data will be used, when in situations where data will be used to make high stakes or consequential decisions (Kane, 2016), training or calibration may be critical; likewise, such approaches play a crucial role in evaluating item alignment to standards (Sireci & Faulkner-Bond, 2014). We encourage others to replicate our approach of including a small set of expertise-linked distractors as a means to assess ownership effects, especially for relatively novel constructs. Such studies could then be

evaluated in aggregate, providing a window into how resilient expert judgement may be.

References

AERA, APA, & NCME. (2014). *The standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Ahearn, L. M. (2001). Language and agency. *Annual Review of Anthropology, 30*(1), 109-137. doi:10.1146/annurev.anthro.30.1.109

Ali, S. H., Carr, P. A., & Ruit, K. G. (2016). Validity and reliability of scores obtained on multiple-choice questions: Why functioning distractors matter. *Journal of the Scholarship of Teaching and Learning, 16*(1), 1-14. doi:10.14434/josotl.v16i1.19106

Avsec, S., & Szewczyk-Zakrzewska, A. (2018). Engineering students' self-efficacy and goal orientations in relation to their engineering design ability. *Global Journal of Engineering Education, 20*(2), 85-90. doi: http://www.wiete.com.au/journals/GJEE/Publish/vol20no2/01-Avsec-S.pdf

Beckstead, J. W. (2009). Content validity is naught. *International Journal of Nursing Studies, 46*(9), 1274-1283. doi:10.1016/j.ijnurstu.2009.04.014

Bonaccorsi, A., Apreda, R., & Fantoni, G. (2020). Expert biases in technology foresight. Why they are a problem and how to mitigate them. *Technological Forecasting and Social Change, 151*(119855), 1-17. doi:10.1016/j.techfore.2019.119855

Bratman, M. E. (2013). *Shared agency: A planning theory of acting together*. New York, NY: Oxford University Press.

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research, 5*(4), 194-197. doi:10.1016/S0897-1897(05)80008-4

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). London, UK: Sage publications.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2016). *Internet, phone, mail and mixed-mode surveys: The tailored design method*. Hoboken, NJ: Wiley.

Ecken, P., Gnatzy, T., & Heiko, A. (2011). Desirability bias in foresight: Consequences for decision quality based on delphi results. *Technological Forecasting and Social Change, 78*(9), 1654-1670. doi:10.1016/j.techfore.2011.05.006

Gable, R., & Wolf, M. (1993). *Instrument development in the affective domain*. New York, NY: Springer Science & Business Media.

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health, 20*(3), 269-274. doi:10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247. doi:10.1037/1040-3590.7.3.238

Holland, D., Lachicotte, W., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.

Howard, M. C. (2018). Scale pretesting. *Practical Assessment, Research & Evaluation, 23*(5), 1-14. doi:10.7275/hwpz-jx61

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319-342. doi:10.1111/j.1745-3984.2001.tb01130.x

Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice, 23*(2), 198-211. doi:10.1080/0969594X.2015.1060192

Kirk, C. P. (2017). When good fences make good customers: Exploring psychological ownership and territoriality in marketing. In C. Olckers, L. van Zyl, & L. van der Vaart (Eds.), *Theoretical orientations and practical applications of psychological ownership* (pp. 135-157). Basel, Switzerland: Springer International Publishing.

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes, 75*(1), 23-55. doi:10.1006/obhd.1998.2781

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563-575. doi:10.1111/j.1744-6570.1975.tb01393.x

Loke, W. H., & Tan, K. F. (1992). Effects of framing and missing information in expert and novice judgment. *Bulletin of the Psychonomic Society, 30*(3), 187-190. doi:10.3758/BF03330437

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382-386. doi:10.1097/00006199-198611000-00017

Martuza, V. R. (1977). *Applying norm-referenced and criterion-referenced measurement in education*. Boston, MA: Allyn & Bacon, Incorporated.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain*. New York, NY: Springer.

Montibeller, G., & Von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis, 35*(7), 1230-1251. doi:10.1111/risa.12360

Morewedge, C. K., Shu, L. L., Gilbert, D. T., & Wilson, T. D. (2009). Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *Journal of Experimental Social Psychology, 45*(4), 947-951. doi:10.1016/j.jesp.2009.05.014

Narayan, D., & Petesch, P. (2007). Agency, opportunity structure, and poverty escapes. In D. Narayan & P. Petesch (Eds.), *Moving out of poverty: Cross-disciplinary perspectives on mobility* (pp. 1-44). Washington, DC: The World Bank.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578. doi:10.3102/00346543066004543

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489-497. doi:10.1002/nur.20147

Richardson, G. M., & Liang, L. L. (2008). The use of inquiry in the development of preservice teacher efficacy in mathematics and science. *Journal of Elementary Science Education, 20*(1), 1-16. doi:10.1007/BF03174699

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94-104. doi:10.1093/swr/27.2.94

Salkind, N. J. (2010). *Encyclopedia of research design* (Vol. 1). Thousand Oaks, CA: Sage.

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema, 26*(1), 100-107. doi:10.7334/psicothema2013.256

Stelly, D. J., & Goldstein, H. W. (2007). Application of content validation methods to broader constructs. In S. M. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validity evidence* (pp. 252-316). San Francisc, CA: Jossey-Bass.

Svihla, V., Gallup, A., & Kang, S. P. (2020). Development and insights from the measure of framing agency. *Proceedings of the 127th ASEE Annual Conference & Exposition*. doi:10.18260/1-2--34442

Svihla, V., Gomez, J. R., Watkins, M. A., & Peele-Eady, T. B. (2019). Characterizing framing agency in design team discourse. *Proceedings of the ASEE 126th Annual Conference and Exhibition*. doi:10.18260/1-2--32505

Svihla, V., & Peele-Eady, T. B. (2020). Framing agency as a lens into constructionist learning. *Proceedings of Constructionism 2020*, 313-324. doi: http://www.constructionismconf.org/wp-content/uploads/2020/05/C2020-Proceedings.pdf

Thom-Santelli, J., Cosley, D., & Gay, G. (2010). What do you know? Experts, novices and territoriality in collaborative systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1685-1694. doi:10.1145/1753326.1753578

Thom-Santelli, J., Cosley, D. R., & Gay, G. (2009). What's mine is mine: Territoriality in collaborative authoring. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1481-1484. doi:10.1145/1518701.1518925

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics, 106*(4), 1039-1061. doi:10.2307/2937956

von der Trenck, A. (2015). "It's mine." The role of psychological ownership and territoriality in knowledge hiding. *Proceedings of the Thirty Fifth International Conference on Information Systems*.

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development, 45*(3), 197-210. doi:10.1177/0748175612440286

Wyse, A. E., & Babcock, B. (2018). A comparison of subject matter experts' perceptions and job analysis surveys. *Practical Assessment, Research, and Evaluation, 23*(1), 1-10. doi:10.7275/7dey-zd62

Citation:

Svihla, V., & Gallup, A. (2021). Do loss aversion and the ownership effect bias content validation procedures? *Practical Assessment, Research & Evaluation*, 26(7). Available online: https://scholarworks.umass.edu/pare/vol26/iss1/7/

Corresponding Author:

Vanessa Svihla
College of University Libraries & Learning Sciences
The University of New Mexico
Albuquerque, NM

Email: vsvihla [at] gmail.com