


COMMENTARY

Structured interviews: moving beyond mean validity . . .

Allen I. Huffcutt¹ and Sara A. Murphy² 

¹University of Wisconsin, Green Bay, WI, USA and ²The University of Winnipeg, Manitoba, Canada

Corresponding author: Allen I. Huffcutt; Email: ahuffcutt60@gmail.com

As interview researchers, we were of course delighted by the focal authors' finding that structured interviews emerged as the predictor with the highest mean validity in their meta-analysis (Sackett et al., 2023, Table 1). Moreover, they found that structured interviews not only provide strong validity but do so while having significantly lower impact on racial groups than other top predictors such as biodata, knowledge, work samples, assessment centers, and GMA (see their Figure 1).

Unfortunately, it also appears that structured interviews have the highest variability in validity (i.e., .42 +/- .24) among top predictors (Sackett et al., 2023; Table 1). Such a level of inconsistency is concerning and warrants closer examination. Given that the vast majority of interview research (including our own) has focused on understanding and improving mean validity as opposed to reducing variability, we advocate for a fundamental shift in focus. Specifically, we call for more research on identifying factors that can induce variability in validity and, subsequently, on finding ways to minimize their influence.

Our commentary will highlight several prominent factors that have the potential to contribute significantly to the inconsistency in validity. We group them according to three major components of the interview process: interview format/methodology, applicant cognitive processes, and contextual factors.

Format/methodology

Utilization of structuring elements

There are as many as 18 different structuring elements, with an average of six being incorporated (Levashina et al., 2014). Unfortunately, both the number and choice of elements varies across studies, representing an important, but largely unexplored, source of validity variance. To illustrate, although rating each question individually using behaviorally anchored rating scales is common, some interview developers use generic scaling (e.g., Likert; Latham & Saari, 1984) and/or wait until the interview is over to provide their ratings (e.g., Kluemper et al., 2015). Whether to allow, discourage, or even prohibit probing is yet another common methodological variation.

Understanding how these structuring elements relate to validity, both individually and interactively, is one of the top directions for future research. In addition to reducing variability in validity, such understanding could encourage utilization of the most effective elements and/or combination of elements and inform best practices in the field.

Wording of structured interview questions

Surprisingly little attention has been paid to how structured interview questions are worded and how that wording affects construct assessment and validity. One such consideration is the

difference between typical (day-to-day tendencies) and maximal performance. Consider the question “Can you tell me about an occasion where you had to deal with an angry client?” (Bangerter et al., 2014), which could easily be answered with an experience reflecting either type of performance. Given the low correlation between typical and maximal behavior, differences in wording could substantially increase validity variance.

We urge interview developers to pay closer attention to the wording of their questions and the resulting effects. Careful identification of the performance type they wish to assess (typical or maximal) should occur upfront. Once identified, questions should be written in a manner that ensures consistent assessment of that type across candidates. For example, a recent study found that “priming” candidates to present past experiences that portray top-end skill capability was in fact associated with more consistent maximal responding (Huffcutt et al., in-press). Consider the question from the Bangerter et al. (2014) study. If maximal performance was the goal, it could be reworded to something like “Can you tell me about an occasion when you were particularly effective in dealing with a really angry client?”

Applicant cognitive processes

Potential for cognitive overload

A growing number of interview researchers are highlighting the potential for cognitive overload in structured interviews. Specifically, the cognitive processes involved in effective responding to some types of questions may be too cumbersome for a number of candidates (Broisy et al., 2020). An average working memory capacity of only four slots (Cowan, 2010), susceptibility to the effects of stress, and frequent use of mental shortcuts (heuristics) all highlight fundamental reasons for such difficulty (see Huffcutt et al., 2020, for a review and discussion).

Consider structured interview questions that ask for description of past experiences. There are no less than five cognitive steps involved, including deciphering the question, searching long-term memory for relevant experiences, comparing/contrasting those experiences to identify the optimal one to present, recalling as many details as possible about that experience, and then formatting for oral delivery (Huffcutt et al., in-press). Common convention is for the interviewer to inquire about all aspects of an experience at once (in a single question), frequently structured around the STAR reporting format (situation, tasks, actions, results),¹ and to expect a unified and complete response shortly after a question is read. Such an undertaking does indeed seem difficult given all of the cognitive limitations noted above, particularly with expectations of a somewhat immediate response.

Although difficulties with cognitive overload emerge at the individual candidate level, there is strong potential for study-to-study variation in both the frequency and degree to which it occurs. In addition to random variation in applicant capability, there may be contexts or even occupations where really difficult challenges do not tend to occur very often, thereby making the search for prime experiences in long-term memory more cumbersome. Conversely, there are areas (e.g., customer service) where difficulties occur quite frequently and are easily recalled. In turn, these variations have the potential to induce variability into resulting validity.

We encourage exploration of alternative approaches that can reduce the cognitive load of structured interview questions. For questions focusing on past behavior, interviewers could employ strategies similar to those used to gather eye-witness testimony in the legal field (which, conceptually, is very similar to describing a past experience). Legal professionals guide the witness through memory recall using a series of small steps, taking their time and building memory recall

¹Some versions of the past behavior response format also encourage candidates to reflect on their experience (i.e., STARR format; Indeed Editorial Team, 2022), adding an additional layer of cognitive complexity.

gradually (see Wade & Spearing, 2023). To illustrate using an IT position, rather than asking candidates to describe all aspects of a situation when they solved a difficult computer issue at once, the interviewer could first prime memory recall by opening with “I am sure you have had to deal with a lot of difficult computer issues during your time as an IT specialist.” A series of follow-up inquiries would then ensue that eventually isolate a single, high-quality experience replete with all of its rich details.

Communication (oratory) ability

One aspect of traditional face-to-face interviewing that has been consistently overlooked is their heavy reliance on oratory communication. This is true of all structured interview types, although the nature and extent of that communication may vary in accordance with format. For example, several researchers have likened description of past experiences to narrative “storytelling” (Bangerter et al., 2014; Brosy et al., 2020), whereas responding to job knowledge questions focuses more on precise conveyance of relevant knowledge (Hartwell et al., 2019).

Some candidates have the natural ability to present themselves, their opinions, their intentions, and their experiences as seeming wonderful, spectacular, and promising regardless of whether they really are so. Even with carefully devised behavioral rating scales, interviewers are not immune from such effects, with ample work finding that impression management tactics are frequent and effective in structured interviews (Ellis et al., 2002).

In jobs where oral communication ability is a core competency (e.g., professional sales, marketing), oratory ability becomes a legitimate aspect of the selection process. However, there are many positions for which it is not a core competency and becomes error variance. Consider a candidate for a tax accountant position who excels in all technical aspects of the field but is perceived as being a little distant and disengaged in the interview. This candidate might be outshone by a candidate who is livelier and more interpersonally engaging but less technically capable. Yet, in the long run, the former candidate could very well contribute more to organizational success.

Unfortunately, much of the research on oratory influences on the interview process has focused on mechanistic aspects such as word count, pitch, rate, and number of pauses (e.g., DeGroot & Motowidlo, 1999) and/or on general constructs such as communication apprehension and fluency. These aspects, although meaningful, do not necessarily capture the full impact that candidate communication has on interviewers and subsequent interview performance ratings. Rather, the impact may be holistic and represent more than the sum of these mechanistic components. We encourage additional research focused on the totality of oratory impact, including inspection of other bodies of literature (e.g., charisma).

Contextual factors

Job type

Type of job is an important moderator that could impact variability in structured interview validity (Sackett et al., 2023). There is empirical evidence suggesting that some forms of structured interviews are less effective for positions of high complexity (Huffcutt et al., 2004). In theory, it should be possible to develop a structured interview for all types of positions regardless of complexity or other considerations. To illustrate, interview developers creating hypothetical scenarios need to ensure that questions and rating scales are sufficiently complex for higher level positions. With past behavior questions, additional aspects or details may need to be gathered for these positions, such as understanding the financial and/or legal position of the company at the time an experience occurred.

Interview etiquette

Interview etiquette (i.e., the customary behaviors undertaken in interview contexts) significantly affects ratings above and beyond the impact of impression management tactics (Tews et al., 2018). Despite their important role in shaping interview interactions, we know little about how these expectations differ across interview contexts and/or the degree to which structure minimizes their influence.

Because etiquette tends to be context dependent (Tews et al., 2018), it is possible that expectations vary according to the job, organization, industry, and/or culture. Differences in the degree to which candidates meet or fail to meet preconceived expectations across studies could, potentially, lead to increased variability in validity. Given that so little is currently known about the role that expectations of etiquette play in structured interviews, we encourage more research on how they exert influence and vary across applicants and situations.

General summary

We may be on the doorstep of a major paradigm shift in interview research, one where the primary focus is on validity variability rather than on mean validity. Careful investigation of factors such as the utilization of structuring elements, wording of questions, potential for cognitive overload, narrative ability, job type/level, and interview etiquette are all excellent avenues to pursue, and there no doubt may be more. Reducing validity variability is, by itself, a worthy goal, but in doing so, there is potential.

References

- Bangerter, A., Corvalan, P., & Cavin, C. (2014). Storytelling in the selection interview? How applicants respond to past behavior questions. *Journal of Business and Psychology*, *29*(4), 593–604. <https://doi.org/10.1007/s10869-014-9350-0>
- Brosy, J., Bangerter, A., & Ribeiro, S. (2020). Encouraging the production of narrative responses to past-behaviour interview questions: Effects of probing and information. *European Journal of Work and Organizational Psychology*, *29*(3), 330–343. <https://doi.org/10.1080/1359432X.2019.1704265>
- Cowan, N. (2010). The magical mystery of four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. <https://doi.org/10.1177/0963721409359277>
- DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*(6), 986–993. <https://doi.org.ezproxy.uwgb.edu/10.1037/0021-9010.84.6.986>
- Ellis, A. P. J., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of Applied Psychology*, *87*(6), 1200–1208. <https://psycnet.apa.org/doi/10.1037/0021-9010.87.6.1200>
- Hartwell, C. J., Johnson, C. D., & Posthuma, R. A. (2019). Are we asking the right questions? Predictive validity comparison of four structured interview question types. *Journal of Business Research*, *100*, 122–129. <http://dx.doi.org.ezproxy.uwgb.edu/10.1016/j.jbusres.2019.03.026>
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Klehe, U. C. (2004). The impact of job complexity and study design on situational and behavior description interview validity. *International Journal of Selection and Assessment*, *12*(3), 262–273. https://doi.org.ezproxy.uwgb.edu/10.1111/j.0965-075X.2004.280_1.x
- Huffcutt, A. I., Howes, S. S., Dustin, S. L., Chmielewski, A. N., Marshall, C. A., Metzger, R. L., Gioia, V. P. (2020). Empirical assessment of typical versus maximal responding in behavior description interviews. *Human Performance*, *33*(5), 447–467. <https://doi.org.ezproxy.uwgb.edu/10.1080/08959285.2020.1812075>
- Huffcutt, A. I., Howes, S. S., Murphy, D. D., & Murphy, S. A. (in press) *Enhancing consistency of maximal responding in behavior description interviews: An exploration of priming and response length*. Personnel Assessment and Decisions.
- Indeed Editorial Team (2022). How to use the STARR method in interview responses. *Indeed*. <https://uk.indeed.com/career-advice/interviewing/starr-method#:~:text=The%20STARR%20method%20is%20an,can%20handle%20real%20work%20scenarios>
- Kluemper, D. H., McLarty, B. D., Bishop, T. R., & Sen, A. (2015). Interviewee selection test and evaluator assessments of general mental ability, emotional intelligence and extraversion: Relationships with structured behavioral and situational interview performance. *Journal of Business and Psychology*, *30*(3), 543–563. <https://psycnet.apa.org/doi/10.1007/s10869-014-9381-6>

- Latham, G. P., & Saari, L. M.** (1984). Do people do what they say? Further studies on the situational interview. *Journal of Applied Psychology*, *69*(4), 569–573. <https://psycnet.apa.org/doi/10.1037/0021-9010.69.4.569>
- Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A.** (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*(1), 241–293. <https://doi.org/10.1111/peps.12052>
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F.** (2023). Revisiting the design of selection systems in light of new findings regarding the validity of widely used predictors. *Industrial and Organizational Psychology: Perspectives on Science and Practice* *16*(3), 283–300
- Tews, M. J., Stafford, K., & Michel, J. W.** (2018). Interview etiquette and hiring outcomes. *International Journal of Selection and Assessment*, *26*(4), 164–175. <https://doi.org/10.1111/ijsa.12228>
- Wade, A., & Spearing, E. R.** (2023). The effect of cross-examination style questions on adult eyewitness accuracy depends on question type and eyewitness confidence. *Memory*, *31*(2), 163–178. <https://doi.org/10.1080/09658211.2022.2129066>