2022

# Interviews from Scratch: Individual Differences in Writing Interview Questions

Lauren J. Wegmeyer
*Wayne State University*

Andrew P. Tenbrink
*Wayne State University*

Angie Y. Delacruz
*Wayne State University*

Rouan Salim
*Wayne State University*

Andrew B. Speer
*Wayne State University*

Follow this and additional works at: https://scholarworks.bgsu.edu/pad

Part of the Human Resources Management Commons, Industrial and Organizational Psychology Commons, and the Other Psychology Commons

How does access to this work benefit you? Let us know!

## Recommended Citation

# INTERVIEWS FROM SCRATCH: INDIVIDUAL DIFFERENCES IN WRITING INTERVIEW QUESTIONS

**Lauren J. Wegmeyer[1], Andrew P. Tenbrink[1], Angie Y. Delacruz[1], Rouan Salim[1], and Andrew B. Speer[1]**

1. Wayne State University

## ABSTRACT

Against best practice recommendations, interviewers prefer unstructured interviews where they are not beholden to regimentation. In cases where interviews are less structured, the interviewer typically generates his or her own set of interview questions. Even in structured interviews though, the initial interview content must be generated by someone. Thus, it is important to understand the different factors that influence what types of questions individuals generate in interview contexts. The current research aims to understand the types of interview questions individuals generate, factors that affect the quality of those questions, how skill in generating interview questions relates to skill in evaluating existing interview questions, and how individual traits relate to skill in generating interview questions. Results show that respondents who are skilled in evaluating existing interview questions are also skilled in writing interview questions from scratch, and these skills relate to general mental ability and social intelligence. Respondents generated questions that most commonly assessed applicant history and self-perceived applicant characteristics, whereas only 30% of questions generated were situational or behavioral.

## KEYWORDS

interview design, social intelligence, question preference, general mental ability

Structured interviews (Huffcutt & Arthur, 1994) are a well-known and highly predictive preemployment assessment of future job performance (e.g., McDaniel et al., 1994; Schmidt & Hunter, 1998). However, despite empirical evidence suggesting that structured interviews are superior to unstructured interviews, unstructured or semistructured interviews are still frequently used in practice (Chapman & Zweig, 2005; Huffcutt & Arthur, 1994; Levashina et al., 2014; Lievens & De Paepe, 2004; Tsai et al., 2016). Interviewers often report preference for unstructured interviews because it allows them to have more informal communication (e.g., "getting to know the applicant") as well as discretion in what questions are asked (Kausel et al., 2016).

Because interviewers continue to use unstructured interviews despite best practice recommendations, research investigating unstructured interviews and how they are developed is important to understand the strategies that are used and possibly how to improve upon them in an unstructured setting. The current research addresses one feature particularly salient to unstructured interviews, though this applies to both unstructured and structured interviews, and that is how interviews are generated—what types of interview questions do individuals generate and what factors affect the quality of generated questions.

Recent studies have found that individual differences are likely to influence what types of questions people prefer when designing interviews (Highhouse et al., 2018; Speer, et al., 2019, 2020; Wright et al., 2012; Zhang, 2021). However, this research has primarily involved participants simply viewing a set of premade interview questions and making judgments regarding those questions. Although useful, this research is less applicable to the common interview process where interviewers *generate* questions from scratch. Thus, the current research examines the actual generation of interview questions from scratch and how individual differences influence this skill. Specifically, this study examined the types of interview questions people generate, how skill in generating interview questions re-

Corresponding author:
Lauren J. Wegmeyer
Author Email: laurenwegmeyer@wayne.edu

lates to skill in evaluating existing interview questions, and how cognitive ability and personality traits relate to skill in generating interview questions from scratch. Because interviewers are frequently conducted by hiring managers with little experience or training on how to conduct interviews, we examined these phenomena using a sample of naïve, untrained respondents.

## Designing Interview Questions

In an unstructured interview setting, new questions may be generated by an interviewer for each applicant or generated once and asked consistently across applicants, regardless of whether they are job relevant or not. In structured interview settings where standardization is high and questions are job relevant, someone still must write the questions from scratch initially. Stated differently, whether a structured or unstructured interview context, the questions that are asked of applicants must at some point be generated by individuals. Though question generation is likely to be particularly prominent for unstructured settings, it is likely relevant when structure is applied as well.

In line with Funder's (1995) realistic accuracy model, good interview questions aim to elicit applicant responses that are reflective of job-related traits (i.e., relevant knowledge, skills, abilities, and other characteristics [KSAOs]; Speer et al., 2020). This requires the interview developer to determine which KSAOs are relevant, either based on a work analysis or a more informal review of the job demands, and then write or select interview questions that would be most effective in eliciting responses related to job-relevant KSAOs.

Research indicates that skill in interview design is relatively stable when evaluating interview questions across different job contexts, such that that interview developers exhibit consistency in identifying good questions across jobs (Speer et al., 2019, 2020). Further, although it is likely that individuals can be trained to generate good interview questions, it has also been shown that individual traits (i.e., general mental ability, social intelligence) are related to skill in question evaluation (Speer et al., 2019, 2020). However, it is unclear whether this skill generalizes to situations where interviewers must create interview questions from scratch (i.e., generation), which more realistically represents the interview context, especially in an unstructured setting. On one hand, both evaluating existing questions and developing questions from scratch require people to understand job demands and which KSAOs should be targeted, which suggests that performance on these two interview design tasks should be consistent. On the other hand, evaluating existing questions is a "recognition" task, whereas generation is more complex. When evaluating existing questions, interview developers must still understand how questions elicit job-relevant KSAO information, but it is likely most developers can easily identify poor questions or recognize

good questions that they have seen previously and evaluate accordingly.

Question generation is likely to differ in several ways. First, different interviewers may adopt different approaches to interview generation. Some may seek to develop questions organically that target relevant KSAOs for the specific job in question, fitting within the theoretical model described previously. On the other hand, memory and previous experiences may influence question generation through the type of information that is readily accessible in an interviewer's memory. For example, when determining what questions to ask an applicant, the interviewer might search their memory for questions that they were asked in previous interviews. Even when trying to target a specific KSAO, an interview developer might rely on past experiences as inspiration (e.g., what they have been asked, what they have asked others in the past, what they have seen others be asked, what they have read). This might suggest that cognitive abilities might be less important when generating questions, as past experiences might obscure the complex process of linking job demands to question prompts in attempts to gather KSAO-related information. Second, and likely counteracting a diminished role of cognitive abilities, skill in developing interview questions requires structuring language in the form of a question. This requires either advanced recall abilities to remember effective question formats that match the situation at hand, or it requires high levels of verbal reasoning to write effective question prompts. Either way, this task likely requires higher levels of cognitive processing.

Given these points, there are reasons to suspect that question evaluation and question generation are distinct but also psychologically and cognitively related. Because these tasks share related processes, those who are better able to evaluate good or effective interview questions should also be better at writing their own interview questions. However, given the differences between evaluation and generation, we would not expect this effect to be large ($\geq .50$).

**Hypothesis 1**: Skill in generating high-quality interview questions is related to skill in evaluating high-quality interview questions.

Skill in evaluating existing interview questions has been shown to be consistent even across jobs that vary greatly in which KSAOs are important (Speer et al., 2020). This is important, as it suggests some people are more capable of customizing questions to elicit job-specific information. Assuming similarity in the underlying psychological processes occurring for question evaluating and question generation:

**Hypothesis 2**: Skill in generating high-quality interview questions is consistent across jobs.

One research question was also proposed around interview design. First, we were interested in understanding the types of interview questions interviewers would generate. Traditional question types include past behavioral or situational. Past behavioral questions elicit responses from applicants pertaining to previous behavior in which they have engaged. Situational questions pose a hypothetical situation to the applicant to which they would respond with how they would handle or behave in that situation. Past research suggests that a wide variety of questions are often asked in unstructured settings, including infrequently researched but commonly used traditional interview questions such as, "Why do you want to work here?" background questions such as, "Tell us about your work history?" and questions asking generally about self-perceived characteristics such as work styles, strengths, and weaknesses (Campion et al., 1997; Hartwell et al., 2019; Posthuma et al., 2014). Little research has examined the prevalence of these different question types, and as such, this research examined the frequency by which these types were generated.

**Research Question 1**: What type of interview questions do people generate?

**Traits Related to Interview Design Skill**

Finally, and as discussed, understanding job demands and how interview questions elicit KSAO-related information linked to those demands is a cognitively demanding task (Speer et al., 2019, 2020). Those with higher general mental ability (GMA) should be better at not only identifying but also writing interview questions from scratch (Morgeson & Dierdorff, 2011). Speer and colleagues (2019, 2020) also found social aptitude (Lievens & Chan, 2010) to correlate with skill in effective interview design. Those with high levels of social aptitude (i.e., socially intelligent, emotionally intelligent, dispositional intelligence) are better able to understand human behavior, how behavior is displayed in the workplace, and how questions might elicit important KSAO-related information in the interview context. Focusing on the effects of these traits on skill in question generation:

**Hypothesis 3**: Skill in generating high-quality interview questions is related to GMA and social aptitude.

## METHODS

**Participants and Procedure**

The sample consisted of 132 undergraduate students who received extra credit in psychology courses for completing an interview simulation task. These participants can be considered naïve respondents with no training in interview design, similar to many hiring managers who conduct interviews in practice and consistent with past interview design studies (e.g., Highhouse et al., 2018; Speer et al., 2019). Most hiring managers are not educated in interviewing and only consider it when the need arises to hire someone.

The sample was 52% White, 67% female, and the average age was 21.29 years. All participants had been interviewed at least once. Participants took a battery of assessments, evaluated a set of existing interview questions, and generated their own questions. Question evaluations and generation were made for two separate jobs (customer service representative [CSR], factory assembler) varying in job duties.

To begin the study, participants first reviewed the job description (JD) for one of the two jobs, which were presented in a counterbalanced order. Job descriptions were adapted from an organization that had existing positions and company specific competencies linked to preferred qualifications for each position. The assembler position had five important job competencies (concern for quality, achieves results, planning and organizing, problem solving, and safety awareness), and the CSR position listed seven important competencies (achieves results, communication, manages conflict, customer focus, stress management, problem solving, and influence and persuasion). Each job-specific competency was listed in the preferred qualifications of the JD and included lists of behaviors that aligned to that competency.

After passing comprehension checks to ensure participants had thoroughly read the JDs, participants were first asked to generate four interview questions from scratch based on what they would ask if they were responsible for conducting interviews to hire for each position. This was the question generation task. After doing this, participants then evaluated 47 existing interview questions on how effectively the preexisting questions measured important applicant characteristics, representing the question evaluation task (Speer et al., 2020). Several types of interview questions were included in this task (e.g., competency specific, traditional, situational, oddball). Competency-aligned questions were obtained from the same organizations in which the job descriptions were obtained.[1] All questions were rated according to "How effective is this question in producing responses that could be used to accurately identify whether an applicant should be hired for the job?" on a scale from 1 (*very ineffective—I would not ask this question*) to 7 (*very effective—I would absolutely ask this question*). Question generation and evaluation were performed for the first job, and then the same procedure was repeated for the other job. Finally, participants took a variety of measures to assess

---

1 The interview questions are available from the first author upon request.

personality and cognitive abilities.

## Question Quality Measures

*Question quality-generation scores and other judgments.* Expert judges rated each of the interview questions generated by respondents regarding several characteristics. Experts were familiar with the interview literature and had written interview questions for previous applied projects. Extensive training and numerous review meetings were conducted to establish a shared frame of reference across the pool of experts (ICC 1,3 agreement = .78 for assembler ratings and .75 for CSR ratings), and three raters rated each respondent and each of the eight questions respondents generated. The raters categorized the questions into various typologies, such as type of question (i.e., past behavioral, situational, self-perceived characteristics, applicant history, administrative, oddball, and/or other), whether it was open ended, leading or loaded, and job relevant. Descriptions for these classifications are in Table 1. Each question was also rated as to the degree it was capable of eliciting information regarding various construct taxonomies taken from Huffcutt et al. (2001) and from the list of company-specific job competencies relevant to the JDs and listed previously. Finally, an overall judgment regarding the effectiveness of each question: "How effective is this question in producing responses that could be used to accurately identify whether

an applicant should be hired for the job?" This latter rating reflects quality of the generated questions, where those who are able to write high-quality interview questions are likely to target job-relevant competencies that would elicit information from an applicant that could be used to identify if the applicant has the needed KSAOs to successfully perform the job. Scores were computed for each question, aggregated into a composite for each job, and then also aggregated across jobs into an overall score, labeled the *question quality-generation*. These scores represent how effectively the participant's generated questions were useful for applicant assessment.

*Question quality-evaluation scores*. To operationalize the quality of the ratings for the existing interview questions, two operationalizations were aggregated in line with procedures conducted by Speer et al. (2020). The first operationalization awarded points if respondents assigned favorable ratings to questions that were linked to job-relevant competencies (identified via work analysis). The second operationalization used expert judgements of the questions and computed a profile correlation between the expert judgments and the participant judgments (ICC 2,4 agreement = .93 for assembler ratings and .89 for customer service representative ratings). These evaluation scores were aggregated into a composite for each position, with high scores

## TABLE 1.

Example Interview Questions by Type

| Question type | Description | Example |
|---|---|---|
| Past behavioral | Asks about past behaviors or accomplishments | Tell us about a time you communicated efficiently and effectively in a difficult situation. |
| Situational | Hypothetical situation posed to applicants | If a customer is very upset about a certain matter, how would you handle that situation? |
| Self-perceived characteristics | Explicit question about personality, job knowledge, strengths, or weaknesses | How well do you work with people? |
| Applicant history | Explicit questions about education, specific work experiences, certifications, or qualifications | What is your level of education? |
| Administrative | Basic logistic of employment, including availability at certain times, travel likelihood, etc. | When would you be able to start and what is your hourly rate preference? |
| Oddball | "Off the wall" questions that cause applicants to "think on their feet" | Why is making products for Company XYZ important and special? |
| Open ended | Allows for higher likelihood of elaboration; unlikely to be a yes or no answer | How would you handle a situation where you witnessed a coworker not following protocol? |
| Leading/loaded | The correct response to the question is obvious | Do you work well with others? |
| Job irrelevant | The question will likely elicit information that is irrelevant to the job | How do you manage your stress? |

*Note.* Examples are generated questions from study respondents.

meaning that participants favorably evaluated job-relevant questions and assigned low ratings to job irrelevant questions. Question quality scores across both jobs were also aggregated into an overall composite, labeled *question quality-evaluation*. Again, this operationalization serves to indicate that respondents were evaluating interview questions positively when they were relevant to each job and likely to elicit job-relevant information about a candidate.

### Individual Difference Measures

*Rater social aptitude assessments.* Two different social aptitude measures were used in this study. The first is the Dispositional Intelligence–Extrapolation (DI-Extr) subfacet measure from the Interpersonal Judgment Inventory (de Kock et al., 2015). Internal consistency was .65. The second social aptitude measure was a work contextualized measure of social intelligence called the Situational Social Intelligence Test (SSIT; Speer et al., 2019). The SSIT is composed of 29 work vignettes in a most and least effective response format (among four possible choices; alpha = .65).

*Rater general mental ability: Wonderlic*. The Wonderlic was used to assess GMA (Wonderlic, 1973). The Wonderlic is a 12-minute timed GMA measure that is commonly used to assess cognitive ability. The Wonderlic typically has acceptable levels of test–retest reliability above .80.

### RESULTS

#### Skill in Question Evaluation Versus Skill in Question Generation

Table 2 reports the descriptive statistics and correlations. Hypothesis 1 examined the relationship between question quality–evaluation scores and question quality–generation scores. The correlation between evaluation and generation questions quality scores was significant for both jobs ($r$ = .26 for assemblers, $r$ = .27 for CSRs). Focusing on composite scores derived as an average across both jobs, question quality–evaluation scores were positively and moderately related to question quality–generation scores ($r$ = .28, $p$ < .05), providing support for Hypothesis 1. Additionally, there was consistency in the quality of generated interview questions across both jobs. Question quality–generation scores from the assembler job were significantly and positively related to the CSR question quality scores ($r$ = .58, $p$ < .01). Thus, Hypothesis 2 was also supported. Like with evaluating the existing interview questions, skill in question generation appears to be consistent even when creating questions for jobs with different demands.

To further examine how questions were generated, mean differences across jobs for each competency were examined and can be found in Table 3. Results from Table 3 suggest that for competencies that were relevant to both jobs, there were smaller differences in competency ratings between jobs than for competencies that were relevant to

only one job (versus the other). For example, for competencies like "achieves results" ($d$ = .32) and "problem solving" ($d$ = -.49), which were relevant within both jobs, there were smaller differences in ratings between jobs. On the other hand, competencies like "customer focus" ($d$ = -1.94), which was only important in one job, differed more substantially in competency ratings between jobs. This suggests that people generally try to customize questions to the specific KSAOs of the job.

#### Trait Relationships With Question Generation Skill

Question quality–generation scores were found to be positively related to GMA ($r$ = .30, $p$ < .01) and social intelligence (SSIT; $r$ = .25, $p$ < .01), supporting Hypothesis 3. However, dispositional intelligence scores were not significantly related to question quality–generation scores (DI-Extr; $r$ = .12, $p$ > .05). As a comparison, relationships between individual trait scores and question quality–evaluation scores appeared to be slightly larger (Table 2). Although the relationships between GMA with question quality–evaluation scores and GMA with question quality – generation scores were not significantly different (Steiger's $z$ = 1.37, $p$ > .05), the relationship between SSIT scores with question quality–evaluation scores was significantly stronger than the relationship between SSIT scores with question quality–generation scores ($z$ = 3.86, $p$ < .01). Thus, these traits generally seem to be more related to the recognition task of evaluating existing questions than skill in generating questions from scratch.

Table 4 reports the frequencies of types of written questions (see Table 1 for question types). Because there was little difference between the frequency of question types across jobs, an overall count was used to summarize results. To facilitate comparisons, we linked question type frequencies from other studies that also investigated how often different question types occur (Hartwell et al., 2019; Posthuma et al., 2014). The Hartwell et al. (2019) study coded interview questions used in a state government agency. The Posthuma et al. (2014) study surveyed employees who had conducted interviews in a diverse sample that included multiple employers and spanned multiple countries.

Perhaps the most interesting question generation finding is that 30% of generated questions were classified as situational or behavioral, which are the two question types most commonly researched in the academic literature. These findings are relatively aligned with Hartwell et al. (2019), in that past behavioral and situational questions were asked less frequently than other types of interview questions. Instead, generated questions in this study most frequently asked about self-perceived characteristics (38%) or applicant history (18%), which is similar to what Hartwell et al. (2019) and Posthuma et al. (2014) found. It should also be noted that some bad interview questions were generated, as 22% of questions were worded as simple yes/no questions

**TABLE 2.**

Descriptive Statistics and Correlations Between Evaluation Question Quality Scores, Generation Question Quality Scores, and Individual Differences

| | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Evaluation CSR Quality Score | 0.00 | 0.90 | - | | | | | | | | |
| 2. Evaluation Assembler Quality Score | 0.00 | 0.94 | .45** | - | | | | | | | |
| 3. Evaluation Composite Quality Score | 0.00 | 0.85 | .85** | .85** | - | | | | | | |
| 4. Generation CSR Quality Score | 3.63 | 0.80 | .27** | .08 | .20* | - | | | | | |
| 5. Generation Assembler Quality Score | 3.11 | 0.74 | .25** | .26** | .30** | .58** | - | | | | |
| 6. Generation Composite Quality Score | 3.37 | 0.69 | .29** | .19* | .28* | .89** | .88** | - | | | |
| 7. GMA | 22.23 | 5.24 | .30** | .44** | .43** | .24** | .30** | .30** | - | | |
| 8. Social intelligence | 8.91 | 8.88 | .44** | .40** | .49** | .22* | .23** | .25** | .39** | - | |
| 9. Dispositional intelligence | 15.39 | 3.39 | .32** | .33** | .39** | .10 | .13 | .12 | .18* | .35** | - |

*Note.* N = 132. * $p < .05$, ** $p < .01$ (two-tailed). Question quality scores are a five-point scale. GMA scores range from 0 to 5. DI-Extr scores range from 0 to 23. SSIT scores range from 0 to 11. GMA = general mental ability. DI-Extra = Dispositional Intelligence – Extrapolation. SSIT = Situational Social Intelligence Test.

**TABLE 3.**

Mean Competency Scores for Generated Questions by Job

| Job competency | Assembler M (SD) | CSR M (SD) | d |
|---|---|---|---|
| Achieves results | **2.04 (.48)** | **1.89 (.43)** | .32* |
| Concern for quality | **1.99 (.49)** | 1.49 (.37) | -.96* |
| Communication | 1.34 (.30) | **2.49 (.53)** | -2.17* |
| Customer focus | 1.11 (.22) | **2.39 (.63)** | -1.94* |
| Influence and persuasion | 1.21 (.36) | **1.80 (.53)** | -1.42* |
| Manages conflict | 1.29 (.30) | **2.25 (.60)** | -1.66* |
| Problem solving | **1.71 (.48)** | **1.99 (.53)** | -.49* |
| Planning and organizing | **1.74 (.46)** | 1.34 (.26) | .87* |
| Safety awareness | **1.69 (.48)** | 1.05 (.17) | 1.24* |
| Stress Management | 1.32 (.38) | **2.15 (.55)** | -1.46* |

*Note.* Scores that are bolded indicate that those competencies were relevant to the job. CSR = customer service representative. d = Cohen's d standardized mean difference. * indicates statistically significant mean difference at $p < .01$.

and 22% were judged to be leading or obvious (e.g., "Are you a hard worker?").

**DISCUSSION**

Understanding how interview questions are constructed is an important area of research given the frequent use of unstructured interviews. Although there is preliminary evidence that suggests individual differences influence the types of interview questions interviewers prefer (e.g., Highhouse et al., 2018; Speer et al., 2019, 2020; Zhang et al., 2021), this research has not considered how individuals *generate* interview questions from scratch. The current study investigated differences in how people generate interview questions, the types of questions generated, and what traits are related to skill in developing good interview questions. We highlight several important findings from this study.

First, this study extends previous work by Speer and colleagues (2020) in showing that skill in interview creation is not limited to evaluating existing questions but also applies to generating new questions. Skill in evaluating existing questions is moderately consistent with effectively generating new questions from scratch. Further, generated questions that asked about job-relevant competencies were specific to the job of interest, similar to findings about the evaluation of existing interview questions from Speer et al. (2020). That said, even though the correlation between scores on these two tasks was positive, it was not an overly

TABLE 4.
Question Type Frequencies

| Question type | Frequency total | Frequency comparison (Hartwell et al., 2019) | Frequency comparison (Posthuma et al., 2014) |
|---|---|---|---|
| Past behavioral | 13% | 7% | |
| Situational | 17% | 5% | |
| Self-perceived characteristics | 38% | | 46% |
| Applicant history (background) | 18% | 56% | 64% |
| Administrative | 1% | | |
| Oddball | 2% | | |
| Other | 11% | | |
| Open-ended question | 78% | | |
| Leading/loaded question | 20% | | |

*Note.* Questions could be categorized as multiple types, although that was not common. Frequency total lists the frequencies of question types in the current study. The two frequency comparison columns list the frequency estimate of the question types in other studies. The Hartwell et al. (2019) study coded interview questions used in a state government agency. The Posthuma et al. (2014) study surveyed employees who had conducted interviews in a diverse sample that included multiple employers and spanned multiple countries.

strong correlation ($r$ = .28). As such, these two tasks do differ, and it is possible that other factors beyond cognitive abilities (as measured here), such as past memory and past interview experiences, may contribute to the generation task. In relation, this study corroborates and extends past research showing that GMA and social intelligence are related to skill in interview design (Speer et al., 2019, 2020). This study showed that this relationship occurs even when people generate questions from scratch. Given these findings, it may be important to focus on these constructs when hiring for positions that interview frequently and especially those in which interviewing is a common task (i.e., recruiters).

Third, this study supports past findings (Posthuma et al., 2014; Hartwell et al., 2019) that show that situational and past behavioral questions, although valid and heavily researched, are not the only types of interview questions likely to be used by interviewers. Instead, the current study

shows that individuals are more likely to generate questions about self-perceived characteristics (38%) and applicant history (18%), as opposed to past behavioral (13%) and situational questions (17%). This is important to note, given that most research in I-O psychology has classified past behavioral and situational question types as the most effective interview questions (Levashina et al., 2014). Clearly, more research should be directed at understanding the less studied question types commonly generated within this study. Furthermore, the aim of the current study was to investigate the types of questions that untrained interviewers generate. What's interesting is that past studies using experienced interviewers (Hartwell et al., 2019) did not report higher usage of behavioral or situational questions, calling into question the role of interviewer expertise. The communalities between the types of questions generated in this study and those generated by more experienced interviewers suggests future research is needed that investigates the role of training and experience when generating interview questions.

Fourth, we found that many generated interview questions were of low quality. Of the questions written, 22% could be answered with a single yes or no response, and 20% had blatantly obvious correct answers (i.e., leading questions). These percentages may be in part due to the lower ability sample used in this study, and it is possible that actual, trained managers and human resources employees would be less likely to write low quality questions. However, these findings also likely highlight the importance for organizations to use some form of interview training to help develop skills in interview creation. That said, training is less likely to occur when unstructured interviews are used. Thus, in situations where training is most needed, training is most likely absent, and therefore academics and practitioners will need to consider other strategies to somehow educate workers on good interviewing practices.

**Limitations and Areas for Future Research**

Given the novelty of this research, there are many study limitations that lend way to areas for future research. First, it is important to highlight that the sample consisted of undergraduate students that are likely less to have experience in interviewing than who would typically generate interview questions. Second and related to the first point, the questions were generated in a simulated context based on job descriptions provided, not in an operational context where the questions are actually used. Thus, results may not generalize when using other samples, and further investigation should use actual interviewers (e.g., trained hiring managers) in organizations. It is likely that experienced interviewers have better awareness of the demands of the job and what KSAOs would be important to target in interview questions as a result. However, in an unstructured setting, interviewers do not typically receive formal training or have education in best practices for conducting interviews.

Likewise, this naïve sample are the types of people who are likely to enter managerial roles, devoid of interviewing experience. Thus, although there are challenges in using a naïve sample, there are also similarities between the sample used and those who would write interview questions.

Third, although we used expert judgments to determine the quality of developed questions, an ideal study might gather additional validity data on the questions that participants generate. Good interview questions should elicit applicant responses that are reflective of job-related traits (i.e., KSAOs), which enables interviewers to assess applicants more accurately. Thus, questions with higher question quality should lead to judgments with higher validity if those questions were to be used. This further highlights the importance of testing these effects in actual applied contexts and focusing on the criterion-related validity of the questions interviewers generate, though such a design might not be feasible in most applied settings.

Last, research suggests that structured interviews are superior to unstructured interviews in part because they are typically based on work analysis information (Huffcutt & Arthur, 1994; Schmidt & Hunter, 1998). Although the format of the questions generated for this study were unstructured (i.e., developed from scratch), respondents viewed the job description thoroughly before writing interview questions. However, in an actual unstructured interview setting, interviewers may or may not reference work analysis materials prior to developing interview questions. Thus, future research should investigate the difference in the types of interview questions that individuals generate when using work analysis materials versus not to fully understand question generation in unstructured settings. Regardless, interview questions must at some point be developed by someone whether the interview is structured or unstructured, and thus we do not expect this study's findings to be restricted to solely unstructured interviews. For example, even if elements of structure are applied (e.g., the same questions are asked across applicants), a poorly generated interview question is likely to affect the interview process.

**Conclusion**

Against best practice recommendations, interviewers typically prefer unstructured interviews where they have the discretion to ask any questions they wish. Because of this, recent research has begun investigating the different factors that influence the design of unstructured interviews. This study contributes to and extends upon the literature by examining how individual differences in interview ability influence the generation of interview questions. In support of previous research, interviewers with higher levels of mental ability and social aptitude are likely to develop better questions. Additionally, this study showed that interview designers will generate a wide range of different types of questions. Most importantly, findings from this study highlight the criticality of interview design training in developing skills for interviewers to generate high quality interview questions.

## REFERENCES

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. Personnel Psychology, 50(3), 655–702.

Chapman, D. S., & Zweig, D. I. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. Personnel Psychology, 58(3), 673–702.

de Kock, F. S., Lievens, F., & Born, M. P. (2015). An in-depth look at dispositional reasoning and interviewer accuracy. Human Performance, 28(3), 199-221.

Funder, D.C. (1995). On the accuracy of personality judgment: A realistic approach. Psychological Review, 102(4), 652-667.

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.

Highhouse, S., Nye, C. D., & Zhang, D. C. (2018). Dark motives and elective use of brainteaser interview questions. Applied Psychology, 68, 311-34.

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. Journal of Applied Psychology, 79(2), 184-190.

Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. Journal of Applied Psychology, 86(5), 897-913.

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. Organizational Behavior and Human Decision Processes, 137, 27-44.

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.

Lievens, F., & Chan, D. (2010). Practical intelligence, emotional intelligence, and social intelligence. In Farr, J.L., & Tippins, N.T., (Eds). Handbook of employee selection (pp. 339-359). Routledge/Taylor & Francis Group.

Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. Journal of Organizational Behavior, 25(1), 29–46.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. Journal of Applied Psychology, 79(4), 599–616.

Morgeson, F. P., & Dierdorff, E. C. (2011). Work analysis: From technique to theory. In S. Zedeck (Ed.), APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization. (pp. 3–41). American Psychological Association.

Posthuma, R. A., Levashina, J., Lievens, F., Schollaert, E., Tsai, W. C., Wagstaff, M. F., & Campion, M. A. (2014). Comparing employment interviews in Latin America with other countries. Journal of Business Research, 67(5), 943-951.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124(2), 262-274.

Speer, A. B., Christiansen, N. D., & Laginess, A. (2019). Social intelligence and interview accuracy: Individual differences in the ability to construct interviews and rate accurately. International Journal of Selection and Assessment, 27(2), 104-128.

Speer, A. B., Tenbrink, A. P., Schwendeman, M. G., & Wegmeyer, L. J. (2020). Individual differences in effective interview design: Factors affecting question choice. International Journal of Selection and Assessment. 28(3), 310-321.

Tsai, W. C., HsinHung Chen, F., Chen, H. Y., & Tseng, K. Y. (2016). When will interviewers be willing to use high-structured job interviews? The role of personality. International Journal of Selection and Assessment, 24(1), 92-105.

Wonderlic, E. F. (1973). Wonderlic personnel test: Manual. Western Psychological Services.

Wright, C. W., Sablynski, C. J., Manson, T. M., & Oshiro, S. (2012). Why are manhole covers round? A laboratory study of reactions to puzzle interviews. Journal of Applied Social Psychology, 42(11), 2834–2857.

Zhang, D. C. (2021). Horse-sized ducks or duck-sized horses? Oddball personality questions are likable (but useless) for organizational recruitment. Journal of Business and Psychology, 1-19.