2021

An Investigation of Interviewer Note Taking in the Field

Jacob S. Fischer
University of Missouri - Saint Louis

James Breaugh
University of Missouri - Saint Louis

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

AN INVESTIGATION OF INTERVIEWER NOTE TAKING IN THE FIELD

Jacob S. Fischer[1] and James Breaugh[1]

1. University of Missouri - Saint Louis

ABSTRACT

KEYWORDS

interviewing, note taking, structured interviews

Although a key component of a structured interview is note taking, relatively few studies have investigated the effects of note taking. To address this lack of research, we conducted a study that examined the effects of note taking in a work setting. As predicted, we found that the total number of notes taken by interviewers and the level of detail of these notes were positively related to the ratings these interviewers gave to job applicants, that interviewer ratings of applicants who were hired were predictive of their job performance ratings, and that interviewer ratings mediated the relationships between note taking and performance ratings (i.e., the number of notes and their level of detail did not have a direct effect on performance ratings). We also showed that, if uncontrolled, interviewer nesting can result in misleading conclusions about the value of taking detailed notes.

Interviewer note taking is an important factor in explaining the effectiveness of structured interviews for selecting employees (e.g., Campion et al., 1997; Roulin et al., 2019). For example, in its guide for conducting structured interviews, the U.S. Office of Personnel Management (2008) stressed that taking "regular and detailed notes" was "crucial" (p. 15). Therefore, it is surprising that interviewer note taking has received relatively little attention from researchers (Levashina et al., 2014). Especially lacking are studies that involved actual interviewers (Blackman, 2017).

In theoretical treatments of note taking, researchers (e.g., Dipboye, 2017; Levashina et al., 2014; Roulin et al., 2019) have suggested that documenting what a job applicant said should positively influence five aspects of the interview process (i.e., interviewer attention, information organization, information storage, interviewer recall, and interviewer judgment). In terms of interviewer attention, note taking has been hypothesized to result in a greater focus on job-related information (Brtek & Motowidlo, 2002). Note taking also has been hypothesized to result in more elaborate processing of the information received from an applicant and this information being stored in a more organized manner in memory (Middendorf & Macan, 2002). Such information storage and being able to review notes should result in an interviewer being better able to recall information obtained during the interview (Burnett et al., 1998). Better recall of job-related information has been hypothesized to result in more accurate interviewer judgments about whether to hire a job applicant (Campion et al., 1997).

Very few studies have tested the theoretical explanations offered for the benefits of note taking. In terms of interviewer attention, Brtek and Motowidlo (2002) showed that students who took more notes while watching videos were rated as being more attentive. Research is lacking with regard to the effects of note taking on how information on a job applicant is organized and is stored in memory. In terms of interviewer recall, a study by Middendorf and Macan (2002) found that observers who took notes while watching videos had better recall of applicant statements than observers who did not take notes. Concerning the accuracy of interviewer judgments, Huffcutt and Woehr (1999) found that in studies in which interviewers took notes there was stronger relationship between interviewer ratings and performance ratings of new employees than in studies in which no notes were taken.

A few researchers (e.g., Burnett et al., 1998) have highlighted the importance of focusing on different aspects of interviewer notes. For example, Brtek and Motowidlo (2002) found that, in comparison to students who took fewer notes while watching an interview video, the interviewer ratings of students who took more notes were more predictive of performance ratings of the employees in the videos made by their supervisors. The level of detail of the notes taken also has been highlighted as meriting attention (Levashina et al., 2014). However, the effects of the level of detail has not been examined.

Corresponding author:
James Breaugh
Email: jbreaugh@umsl.edu

Based on the research reviewed, a few tentative conclusions seem warranted. Note taking appears to result in individuals paying greater attention to what applicants said (Brtek & Motowidlo, 2002) and in more accurate recall of their responses (Macan & Dipboye, 1994). Taking notes (Huffcutt & Woehr, 1999) and taking more notes (Brtek & Motowidlo, 2002) seem to improve the accuracy of interviewer judgments in predicting performance ratings. We characterize these conclusions as being tentative because relatively few studies of note taking have been conducted. Given the hypothesized importance of interviewer note taking, it is understandable why researchers (e.g., Levashina et al., 2014) have stressed the need for additional research that further explores the value of note taking.

Hypothesis Development: Note Taking, Interviewer Ratings, and Job Performance Ratings

Our study addressed the calls for future research (e.g., Dipboye, 2017, Levashina et al., 2014) on interviewer note taking. Using a sample of job applicants who were hired for an administrative assistant position, we focused on the total number of notes taken by an interviewer on an applicant and the level of detail of these notes (as assessed by trained raters). We focused on these two aspects of note taking because both theoretical treatments of note taking and the results of studies suggest their importance. Figure 1 portrays the hypothesized relationships we tested among the total number of notes taken by an interviewer, the detail of these notes, interviewer ratings of job applicants, and supervisor performance ratings.

Hypotheses 1a and 1b concern the relationship between the number of notes taken and the detail of these notes and interviewer ratings of job applicants who were subsequently hired. Theory (e.g., Levashina et al., 2014) suggests that, in comparison to interviewers who have taken fewer notes and less detailed notes, interviewers who have taken more notes and more detailed notes should have focused more heavily on job-related information during an interview and better organized and stored this information in memory. For interviewers who have taken more notes and more detailed notes, having access to this information means they have more job-related information upon which to base an evaluation of a job applicant. Possessing such information should increase an interviewer's confidence because the interviewer has documentation to support his or her rating of an applicant. This confidence should result in an interviewer being more willing to rate an applicant as being either an excellent or a poor candidate for a job opening. More extreme interviewer ratings are important because they help decision makers differentiate among applicants (Huffcutt, 2020 has provided a detailed treatment of problems resulting from range restriction in interview studies). Hypotheses 1a and 1b are stated in terms of a positive relationship rather than a curvilinear one (e.g., that having more detailed notes

should result in a lower or higher interviewer rating) because applicants who received low ratings would not have been hired. In addition to theorizing on the effects of note taking supporting Hypotheses 1a and 1b, a study by Azizi (2015) provides tangentially related empirical evidence that supports these hypotheses. Azizi had students study vignettes of workers that included information on their job performance. Students who were more comfortable with the rating process gave more extreme performance ratings to high and low performing workers. Students who expressed less comfort with the rating process tended to use the middle of the performance rating scale. With regard to Hypotheses 1a and 1b, assuming that having taken more notes and more detailed notes results in greater interviewer comfort in rating job applicants, we would expect more extreme interview scores.

> **Hypothesis 1a:** There will be a positive relationship between the number of notes an interviewer takes on a job applicant and the interviewer rating the applicant receives.

> **Hypothesis 1b:** There will be a positive relationship between the level of detail of the notes an interviewer takes and the interviewer rating the applicant receives.

Assuming the notes taken by an interviewer are job related, the number and the detail of these notes should be indirectly associated with performance ratings of job applicants who were hired. This indirect effect is best understood if the influence of note taking on the interviewer rating is considered. As an interviewer takes notes, the interviewer is forming an impression of an applicant (Dipboye, 2017). At the end of the interview, this impression may be influenced by a review of the notes taken (Campion et al., 1997). As reflected in Figure 1, the sequence described suggests the association between an interviewer's notes on an individual and the individual's performance rating should be mediated by the interviewer's rating of the person (i.e., the relationships described in Hypotheses 2a and 2b are indirect effects). As an example of what we are proposing, consider the following situation. During an interview, a job applicant responds to a question in a way that shows a high level of relevant work experience. An interviewer documents this experience in one or more notes. At the end of the interview, these notes result in the interviewer giving the applicant a higher rating than would have been given if the applicant had responded in a manner that showed a lower level of relevant experience. In turn, the interviewer rating of the applicant should be positively associated with the person's performance rating. The sequence described suggests the association between an interviewer's notes on an individual and the individual's performance rating should be mediated by the interviewer's rating of the person.

Hypothesis 2a: Interviewer ratings will mediate the effect of the total number of notes taken on job performance ratings.

Hypothesis 2b: Interviewer ratings will mediate the effect of the level of detail of the notes taken on job performance ratings.

The mediated relationships described in Hypotheses 2a and 2b are based on an assumption that interviewer ratings predict performance ratings. Although we examined this relationship (Hypothesis 3), given it has been well-established (Levashina et al., 2014), it seems unnecessary to provide a detailed rationale for it. Rather, it should suffice to state that it is generally assumed that an interviewer gathers information on job applicant variables that should predict job performance (Blackman, 2017).

Hypothesis 3: There will be a positive association between interviewer ratings and job performance ratings.

A Research Question

In conducting a study, researchers often have ignored the possibility of an interviewer nesting effect even though it could confound the hypothesized relationships being tested (Hartwell & Campion, 2016). In multilevel analysis terms, a nesting effect refers the relationship between two level-1 variables not being independent of a level-2 variable. For example, in our study, it is possible that some interviewers may take more notes and more detailed notes and also give higher interviewer ratings. The failure to control for interviewer effects can result in erroneous conclusions (Scherbaum & Pesner, 2019). Given this possibility, we examined whether the results for our hypotheses having controlled for interviewer effects were different than if we had ignored nesting. To control for nesting, for each interviewer, interviewer ratings were standardized so that the mean interviewer rating was 0.00, and the standard deviation was 1.00.

Research Question: Does controlling for nesting matter in testing our hypotheses?

We believe our study makes four important contributions. First, we investigated the effects of the number of notes and their level of detail on interviewer ratings. Second, we tested whether the relationships between the notes taken and job performance ratings were mediated by interviewer ratings. Third, we investigated whether controlling for interviewer nesting made a difference in terms of what we would conclude about note taking. Finally, our study involved data gathered from an actual work setting in which high stakes decisions were being made.

METHOD

Participants in This Study and the Hiring Process

To test our hypotheses, we needed access to interviewers' ratings of job applicants, notes taken on them by the interviewers, and job performance ratings of these applicants. A financial services firm in the United States allowed us access to these data. The job focused on in our study is an administrative assistant position that involved clerical tasks, interacting with clients, and helping with marketing activities. Among the qualities needed for being an administrative assistant are attention to detail, being organized, and the ability to multitask, problem solve, and communicate effectively. This job is an entry-level, nonexempt position.

As described by a human resource professional from

FIGURE 1.
Hypothesized Model
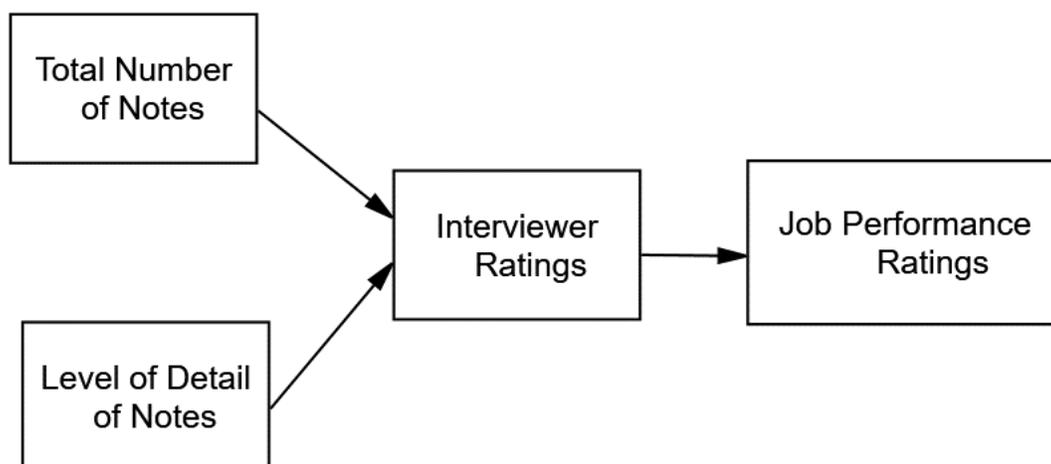
the organization that supplied our data, the selection process for the administrative assistant position began with individuals submitting job applications to the firm's website for an advertised opening. These applications were screened by a corporate interviewer in terms of a person's suitability for the job opening. In making this assessment, key considerations were an applicant's work history (e.g., a history of job hopping would eliminate a person from consideration) and required salary (e.g., a person who stated a required a salary well in excess of that being offered by the organization was seen as a poor fit for the job). If an applicant was viewed positively, a phone interview was conducted by the interviewer. In addition to asking questions of an applicant, information about the administrative assistant position was provided during this interview. For applicants who remained interested in the position after the interview, their applications were forwarded to the investment advisor with the job opening along with an interviewer's ratings of the applicants and the notes taken during the interviews. After reviewing this information, an investment advisor decided which applicant or applicants he or she would interview. Ultimately, an investment advisor decided whom to hire.

For the 6-month period included in our study, the organization had interviewer ratings of applicants, interviewer notes, and job performance ratings on 282 individuals who were hired for the administrative assistant position. These individuals were interviewed by one of 11 corporate interviewers. This sample was reduced to 247 individuals due to two job applicants being dropped because they had interviewer ratings outside of the values specified on the rating scale and 33 applicants being dropped because the individual who interviewed them gave almost identical ratings to everyone. More specifically, across 165 interview ratings (rating five competencies across 33 applicants), this interviewer gave a score of 2 (on the 1–3 interview rating scale) 155 times (94%). This interviewer differed from the other interviewers as a group. Specifically, this outlier interviewer's average interview score was lower than that of the other interviewers ($p < .001$), and this outlier interviewer differed from the other 10 interviewers as a group in terms of taking fewer notes ($p < .001$) and less detailed notes ($p = .02$). Given the applicants interviewed by this interviewer did not differ from those interviewed by the other interviewers in terms of the average performance rating they received ($p = .97$), it is not likely that this outlier interviewer's lower average interview rating and lower scores for total number of notes taken and their level of detail were due to him or her having interviewed a weaker set of applicants.

Measures and Sources of Data

From the host organization, we sought data on background characteristics for the interviewers and job applicants. It was unwilling to share this information due to privacy concerns.

Interviewer Ratings of Job Applicants. The interviewer position in this study was salaried and entry level. For each administrative assistant job opening an interviewer was responsible for filling, the relevant investment advisor provided feedback to the interviewer's manager and the interviewer concerning the subsequent performance of the job applicant who was hired and concerning the quality of the interviewer's performance during the hiring process. Prior to having responsibility for filling job openings, newly hired interviewers were trained. This training involved becoming knowledgeable concerning the administrative assistant position (e.g., duties, needed abilities), becoming familiar with the interviewing form used, learning how to take notes during the interview, mastering telephone interviewing skills, and learning how to work effectively with the investment advisor with a job opening. This training involved mastering a training manual, observation (e.g., watching an experienced interviewer screen resumes, conduct phone interviews, take notes, and rate job applicants), and individual mentoring by an experienced interviewer.

Interviewer ratings were made at the end of a phone interview that typically lasted 20–30 minutes. An interviewer asked behavior-oriented questions that tapped five competencies (i.e., ability to build relationships, confidence in one's ability, ability to multitask, ability to problem solve, and attention to detail) the organization saw as important for success as an administrative assistant based on a job analysis. Among the questions asked were: "Give me an example of a time when you were working on multiple tasks under a deadline?" and "Describe a time when you had to make an important or challenging decision?" At the end of the interview, an interviewer rated an applicant's responses for each of the competencies on a three-point scale with the anchors *not qualified*, *qualified*, and *highly qualified*. Each scale point had behavioral anchors appropriate to the competency assessed. Although interviewers were provided with a 3-point rating scale, a number of them gave half-point ratings (e.g., 2.5). The interviewer score used in this study was the average score an applicant received for his or her responses to questions tapping the five competencies. This interview would best be described as semi-structured (Huffcutt et al., 2014). For example, it included some of the components of a structured interview (e.g., the same questions being asked, notes being taken) discussed by Levashina et al. (2014), but it did not include such components as the control of ancillary information and the use of multiple interviewers for each applicant. The coefficient alpha for the interviewer ratings was .57.

The Notes Taken by Interviewers. For the five competencies assessed, interviewers were expected to take notes on relevant job applicant comments. To facilitate note taking, the interviewer rating form included spaces for notes. During the training they received, interviewers were shown how to take notes and informed their notes would be shared with the investment advisor with the job opening.

To compute the total number of notes taken on an applicant and assess their level of detail, two PhD students, who were blind to our hypotheses, reviewed the interviewer forms for the 247 administrative assistants. Prior to coding notes, the students were instructed that a distinct note could reflect a separate sentence or thought or a separate bullet point in the note-taking text area of the interviewer rating form, and they practiced coding notes that were similar to those on the interviewer rating forms and received feedback from the lead author on their coding. The coders computed the total number of notes taken on an applicant and they rated the overall level of detail of the notes by responding to the following item: "How detailed/extensive are the interview notes on this candidate?" (1 = *very little detail* . . . 3 = *moderate detail* . . . 5 = *extensive detail*). When the coders disagreed, the lead researcher broke the tie.

Supervisor Job Performance Ratings. Six months after hiring, an administrative assistant's performance was rated by the investment advisor to whom the assistant reported. This rating was part of the organization's formal review process. Job performance was rated with a single item that was responded to using a 4-point scale (1 = *below expectations* . . . 4 = *outstanding*).

RESULTS

Because the interviewer ratings and notes came from 10 interviewers, we checked whether scores for these variables were related to the interviewers (ignoring nesting can result in inaccurate conclusions). The intraclass correlations were .34 for the interviewer ratings, .76 for the total number of notes, .62 for their level of detail, and .02 for the performance ratings. These values show the need to control for nesting for variables linked to the interviewers. Because multilevel modeling is inappropriate with only 10 interviewers (Scherbaum & Pesner, 2019), we addressed nesting

by standardizing variables within interviewers. This strategy has been used in a few interview studies for dealing with nesting (Hartwell & Campion, 2016).

In terms of the coding of notes, interrater reliability was .82 for the total number of notes and .77 for their level of detail. Table 1 presents information on the main variables in our study. For interviewer ratings, total number of notes, and the level of detail of the notes, we present information for the original (i.e., unstandardized) and the standardized variables. In terms of original interviewer ratings, the mean value was 2.53 on the 3-point scale, which suggests that applicants who received low interview scores were not hired.

Review of Results

To control for interviewer nesting effects, standardized variables were used in our analyses. We tested our hypotheses using Amos (version 26) with bootstrapping (2,000 iterations) used to generate unbiased 95% confidence intervals. Hypotheses 1a predicted a positive relationship between the number of notes an interviewer took on a job applicant and the interviewer rating the applicant received. Hypothesis 1b predicted a positive relationship between the level of detail of the notes and the interviewer rating. As shown in Table 2, Hypothesis 1a ($\beta = .12$, $p = .05$) and Hypothesis 1b ($\beta = .18$, $p = .01$) received support. The two rightmost columns in Table 2 present results for the original measured variables. These results are discussed when addressing the research question we investigated.

Hypotheses 2a and 2b predicted that interviewer ratings would mediate the effects of the number of notes taken and their level of detail on job performance ratings. As reported in Table 2, the bootstrapped confidence intervals for the indirect effects for the number of notes ($\beta = .02$, $p = .04$) and their level of detail ($\beta = .03$, $p = .02$) on performance ratings do not include 0.00, supporting Hypotheses 2a and 2b. To examine whether the number of notes and their detail had direct effects on performance ratings, we added these direct

TABLE 1.
Means, Standard Deviations, and Correlations of Major Study Variables

Variable	M	SD	1	2	3	4	5	6	7
1. Job performance rating	3.02	.79	--						
2. Interviewer rating	2.53	.29	.15**	--					
3. Total number of notes	13.92	3.37	.14*	.33**	--				
4. Notes level of detail	3.02	1.00	.14*	.23**	.71**	--			
5. Standardized interviewer rating	.00	.99	.15**	.79**	.10	.13*	--		
6. Standardized total # of notes	.00	.98	.05	.15**	.46**	.20**	.17**	--	
7. Standardized notes level of detail	.00	.99	.06	.16**	.17**	.59**	.22**	.28**	--

*Note. N = 247. *p < .05, **p < .01.*

TABLE 2.
Tests of Hypotheses

Hypothesis	Relationship	Effect	β (standardized variables)	Upper/lower confidence interval limits	β (original variables)	Upper/lower confidence interval limits
1a	Number of notes → Interviewer rating	Direct	.12*	.01/.24	.29**	.13/.44
1b	Detail of notes → Interviewer rating	Direct	.18**	.06/.31	.05	-.10/.21
2a	Number of notes → Performance rating	Indirect	.02*	.01/.05	.05*	.01/.10
2b	Detail of notes → Performance rating	Indirect	.03*	.01/.06	.00	-.01/.04
3	Interviewer rating → Performance rating	Direct	.15*	.04/.24	.17**	.06/.27

Note. Indirect effects are mediated by interviewer rating. We tested our hypotheses using Amos (version 26) with bootstrapping (2000 iterations) used to generate unbiased 95% confidence intervals. Exact p-values are provided in the text. *$p < .05$, **$p < .01$

effects to the model in Figure 1. The two note-taking variables did not have direct effects (both β's = .02, both *p*'s = .74)

Hypothesis 3 predicted that interviewer ratings of administrative assistants would be positively associated with their performance ratings. The standardized regression weight in Table 2 for this relationship (β = .15, *p* = .02) is consistent with this hypothesis.

Our research question addressed whether controlling for nesting affected the results for our hypotheses. Using standardized variables, all five of our hypotheses received support. Using the variables as originally measured (no standardization due to nesting), we found support for three hypotheses (see the two rightmost columns in Table 2). The number of notes had a direct effect on interviewer ratings (β = .29, *p* = .01), the number of notes had an indirect effect on performance ratings ((β = .05, *p* = .01), and the interviewer ratings had a direct effect on performance ratings (β = .17, *p* = .01). In contrast to when standardized variables were used, as originally measured, the level of the detail of the notes did not have a direct effect on interviewer ratings (β = .05, *p* = .57) and did not have an indirect effect on performance ratings (β = .00, *p* = 48).

DISCUSSION

Although a key component of a structured interview is note taking, relatively few studies have investigated the effects of note taking, and most of these studies have involved students who watched videos of simulated interviews (Levashina et al., 2014). Given the lack of research conducted with real interviewers in high stakes situations, we conducted a study that examined the effects of note taking in an actual hiring context.

Discussion of Results

Based on prior theorizing concerning the effects of note taking (e.g., Levashina et al., 2014), we hypothesized that taking more notes and more detailed notes on job applicants would result in more positive interviewer ratings. We found support for these hypotheses. In interpreting these results, a key concern is whether the more positive interviewer

ratings are linked to subsequent performance ratings or only reflect that note-taking results in inflated interviewer ratings. We believe the former interpretation makes more sense when considered in light of our results for Hypothesis 3. If higher interviewer ratings simply reflected leniency error, interviewer ratings should not be a valid predictor of performance ratings.

Hypotheses 2a and 2b predicted that the number of notes taken and their level of detail would have indirect effects on job performance ratings through interviewer ratings. Both hypotheses were supported. In considering these findings, it is important to recall that we did not find direct effects for the number of notes and their level of detail on performance ratings. If we had, this could suggest that higher performing individuals generated more notes and more detailed notes. However, the fact that our note-taking variables were not directly linked to performance ratings suggests that applicants with more notes and more detailed notes were not higher quality job candidates, at least as reflected by their performance ratings.

We found a positive relationship between interviewer ratings and job performance ratings. In considering this finding, two factors merit consideration. First, the interviewer ratings and the performance ratings came from different sources (in some studies, the hiring manager provides both ratings, which could result in same source bias confounding the relationship). Second, we used performance ratings supplied by the organization, which reflected supervisor judgments used for administrative purposes. As discussed by Murphy et al. (2018), administrative ratings are prone to leniency error and range restriction (e.g., individuals performing poorly may not be on the job long enough to receive a performance rating), both of which could have affected the interviewer rating validity coefficient.

With regard to our research question, the results for the use of the standardized versus the original variables show the importance of considering interviewer nesting effects. Using the standardized variables, we found support for all five of our hypotheses. When we used the original variables, the level of detail of the notes was not linked directly to the interviewer ratings, and it did not have an indirect effect on the performance ratings. In summary, if we had not

used standardized measures, we might have concluded that taking detailed notes was unnecessary. This inconsistency in results is likely be due to the difference in the size of the correlation between the number of notes and their level of detail when using standardized variables ($r = .28$) versus the original variables ($r = 71$).

Practical Implications

The results of our study show that taking more notes and more detailed notes were linked to higher interviewer ratings, which were in turn linked to higher performance ratings. For hiring managers, having access to some interviewer ratings that are near the top of the rating scale makes it easier to differentiate among applicants in deciding who should receive a job offer.

Although the results we reported for our hypotheses were modest in terms of their magnitude, several factors (e.g., range restriction, one-item measures) may have attenuated the size of the relationships we reported. In this regard, Huffcutt et al. (2014) have shown that, corrected for artifacts, an interviewer validity coefficient can be two or three times larger than the uncorrected coefficient. It seems likely that the same would hold true for the magnitude of the results we reported. For example, if we corrected the correlation of .15 between interviewer ratings and job performance ratings for unreliability using coefficient alpha for the interviewer ratings and an estimate of .70 for the performance measure (Wanous & Hudy, 2001), the corrected correlation is .24. Because we did not have access to the unrestricted standard deviation for all job applicants for the interviewer ratings, we could not correct for range restriction. We would also point out that even small effects can be important in the context of hiring decisions. For example, Hurtz and Donovan (2000) estimated the uncorrected correlations between the Big Five personality traits with job performance ranged from .04 to .14. Despite the magnitude of these validity coefficients for personality traits, they are commonly used for making selection decisions.

When the findings of our study are combined with the fact that taking job-related notes should reduce interviewer bias, increase the confidence of a hiring manager in an interviewer's judgment, and send a positive message to job applicants that attention is being paid to their comments, a strong case can be made for the practical consequences of interviewer note taking.

Study Limitations and Future Research

As with most studies conducted in actual work settings, our study has limitations. For example, our performance measure involved a single rating of job performance. Another limitation was our inability to examine whether the type of notes taken moderated the validity of the interviewer ratings in predicting performance. We did not test for note taking as a moderator because the lack of substantial variation on the interviewer ratings and performance ratings limits our ability to fairly test for moderation (Murphy & Russell, 2017).

In addition to future studies addressing the limitations acknowledged, we would highlight four areas for research. First, gathering information on background characteristics of the job applicants and interviewers would be valuable. This information would allow a consideration of such questions as whether interviewers who take more notes are less likely to exhibit bias toward members of underrepresented groups. A second area meriting attention is an examination of the valence of the notes taken. As with past studies (e.g., Middendorf & Macan, 2002), we did not consider whether the tone of each note was positive or negative. Although it is likely that the majority of the notes taken in our study were positive (otherwise individuals would not have been hired), giving attention to the valence of notes should result in a better understanding of the effects of note taking (e.g., do negative notes carry more weight in a hiring decision than positive notes?). A third area we would like to see pursued in future research is a consideration of the job relatedness of the notes taken and the relative impact of job-related notes versus notes that are not job related. The final issue we would raise as meriting future research is the causal flow of the effects of note taking. In Figure 1, we hypothesized that the number of notes taken and their level of detail affect the interviewer rating. However, it is possible that the causal flow is reversed. That is, an interviewer who feels quite positive about a job applicant at the end of the interview may be more likely to write more notes and more detailed notes once the interview is completed. In our study, we could not disentangle the causal direction of whether notes influence the interviewer rating versus the interviewer rating influences the notes that are taken. Therefore, future research that sheds light on this issue would be beneficial. If such a study was conducted with real interviewers, they could be asked how they viewed the interview process (e.g., did they primarily take notes to justify a positive interviewer rating or primarily let the notes they took result in a given interviewer rating?). Clearly, it is possible that the process is more interactive, such that notes are taken during the interview, an interviewer rating is made, and then more notes are taken.

Conclusion

Although the benefits of interviewers taking job-related notes has been emphasized, these benefits have rarely been shown in actual work settings. The results of our study support the value of note taking. In particular, our results show that having more notes and more detailed notes are linked to higher interviewer ratings, which are linked to higher job performance ratings. Because having interviewers take notes involves little time and no financial costs, asking for such documentation would seem to make sense for most employers.

REFERENCES

Azizi, M. (2015). Does comfort matters? The role of rater discomfort on performance ratings. Dissertation Abstracts International Section A: Humanities and Social Sciences, 75(10-A)(E).

Blackman, M. (2017). Using interviewing in selection. In H. W. Goldsein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), The Wiley-Blackwell handbook of the psychology of recruitment, selection, and employee retention (pp. 182-201). John Wiley & Sons Ltd.

Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. Journal of Applied Psychology, 87(1), 185-191. http://doi:10.1037/0021-9010.87.1.185

Burnett, J. R., Fan, C., Motowidlo, S. J., & DeGroot, T. (1998). Interview notes and validity. Personnel Psychology, 51, 375–396. http://doi:10.1111/j.1744-6570.1998.tb00730.x

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. Personnel Psychology, 50, 655–702. http://doi:10.1111/j.1744-6570.1997.tb00709.x

Dipboye, R. L. (2017). The selection/recruitment interview: Core processes and contexts. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), The Blackwell handbook of personnel selection (pp. 119-142), Blackwell Publishing. https://doi.org/10.1002/9781405164221.ch6

Hartwell, C. J., & Campion, M. A. (2016). Getting on the same page: The effects of normative feedback interventions on structured interview ratings. Journal of Applied Psychology, 101, 757-778. http://dx.doi.org/10.1037/apl0000099

Huffcutt, A.I. (2020). Range restriction in employment interviews: An influence too big to ignore. In E. F. Stone-Romero & P. J. Rosopa (Eds.), Research methods in human resource management: Toward valid research inferences (pp. 173-195). Information Age Publishing.

Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. International Journal of Selection and Assessment, 22(3), 297-309. http://doi:10.1111/ijsa.12078

Huffcutt, A. I., & Woehr, D. J. (1999). Further analysis of employment interview validity: A quantitative evaluation of interviewer-related structuring methods. Journal of Organizational Behavior, 20, 549–560. http://dx.doi.org/10.1002/(SICI)1099-1379(199907)20:4_549::AIDJOB921_3.0.CO;2-Q

Hurtz, G. H., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. Journal of Applied Psychology, 85, 869-879. http://doi: 10.1037//0021-9010.85.6.869.

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, 241–293. http://dx.doi:10.1111/peps.12052

Macan, T. H., & Dipboye, R. L. (1994). The effects of the application on processing of information from the employment interview. Journal of Applied Social Psychology, 24, 1291–1314. http://doi:10.1111/j.1559-1816.1994.tb00559.x

Middendorf, C. H., & Macan, T. H. (2002). Note-taking in the employment interview: Effects on recall and judgments. Journal of Applied Psychology, 87, 293–303. http://doi:10.1037//0021-9010.87.2.293

Murphy, K. R., Cleveland, J. N., & Hanscom, M. (2018). Performance appraisal and Management. Thousand Oaks, CA: Sage.

Murphy, K. R., & Russell, C. J. (2017). Mend it or end it: Redirecting the search for interactions in the organizational sciences. Organizational Research Methods, 20, 549-573. http://doi: 10.1177/1094428115625322

Roulin, N., Bourdage, J. S., & Wingate, T. G. (2019). Who is conducting "better" employment interviews: Antecedents of structured interview component use. Personnel Assessment and Decisions, 5, 37-48. https://scholarworks.bgsu.edu/pad/vol5/iss1/2

Scherbaum, C. A., & Pesner, E. (2019). Power analysis for multilevel research. In S. E. Humphrey & J. M. LeBreton (Eds.), The handbook of multilevel theory, measurement, and analysis. American Psychological Association. http://dx.doi.org/10.1037/0000115-001

United States Office of Personnel Management. (2008). Structured interviews: A practical guide. https://www.opm.gov/policy-data-oversight/assessment-and-selection/structured-interviews/guide.pdf

Wanous, J. P., & Hudy, M. J. (2001). Single-item reliability: A replication and extension. Organizational Research Methods, 4, 361-375. DOI: 10.1177/109442810144003