Belief in the Unstructured Interview: The Persistence of an Illusion

Jason Dana
Dept. of Psychology
University of Pennsylvania[*]

Robyn M. Dawes
Nathanial R. Peterson
Dept. of Social and Decision Sciences
Carnegie Mellon University

This version: August 15, 2012

*Address correspondence to:
Jason Dana
University of Pennsylvania
3720 Walnut St.
Philadelphia PA 19104
danajd@psych.upenn.edu

Dawes was thankful for support from NSF Grant SES-0136259.

Abstract

Unstructured interviews are a ubiquitous tool for making screening decisions despite vast evidence of their invalidity. In three studies, we investigated the propensity for "sensemaking" - the ability for interviewers to make sense of virtually anything the interviewee says – and “dilution” – the tendency for non-diagnostic information to weaken the predictive value of quality information. In study 1, participants predicted two fellow students’ semester GPAs from background information and, for one of them, an unstructured interview. In one condition, the interviewee secretly answered questions according to a random system. Consistent with sensemaking, random interviews did not perturb predictions or diminish perceptions of the quality of information that the interview yielded. Consistent with dilution, participants made better predictions about students whom they did not interview. Study 2 showed that merely watching a random interview, rather than conducting it, did little to mitigate sensemaking. Study 3 showed that participants believe unstructured interviews will help accuracy, so much so that they would rather have random interviews than no interview. Impressions formed from unstructured interviews can seem valid and inspire confidence even when interviews are useless. Our simple recommendation for those making screening decisions is not to use them.

Belief in the Unstructured Interview: The Persistence of an Illusion

In 1979, an act of legislature suddenly forced the University of Texas Medical School at Houston to admit 50 more applicants. The additional applicants were initially rejected for admission, based largely on impressions from unstructured interviews in which each interviewer could ask different questions of different applicants. Apparently, the expense of having faculty interview every applicant was wasted: at the conclusion of medical training and one postgraduate year, there were no meaningful differences between the initially rejected and initially accepted students in terms of attrition, academic performance, clinical performance, or honors earned (Devaul et al., 1987). Several field studies have provided similar examples of the embarrassingly poor validity of unstructured interviews for screening decisions (e.g., Bloom & Brundage, 1947; Milstein, Wilkinson, Burrow, & Kessen, 1981; Carroll, Wiener, Coates, Galegher, & Alibrio, 1982). More systematic reviews in the area of employment decisions likewise show that unstructured interviews are poor predictors of job performance, with structured interviews faring somewhat better (Wiesner & Cronshaw, 1988; Wright, Lichtenfels, & Pursell, 1989; Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994).

Despite the evidence, unstructured interviews remain a ubiquitous and even predominant tool for many screening decisions. Studies of human resource executives suggest that they believe most in the validity of unstructured interviews, even when they are aware that the evidence suggests that structured assessment is superior (Highhouse, 2008). Academics, though not professional interviewers, may decide to accept graduate students or hire faculty based on an informal 20 minute chat, countermanding substantial

aggregated and/or statisticized data comparing the candidate to others (test scores and GPAs in the case of students, C.V.s in the case of faculty). Recently, Wake Forest University stopped requiring standardized tests for undergraduate admissions, moving to a system in which every applicant is eligible for an unstructured interview that figures into the admissions decision in a "holistic," non-numeric manner (Highhouse & Kostek, 2010).

Why do people persist in relying on unstructured interviews despite the evidence? We focus on two possible reasons. First, we are aware of no evidence that unstructured interviews *hurt* accuracy by way of studying the same decision makers with and without access to interviews. Indeed, this point was made without rebuttal in a discussion of the Wake Forest decision on the Society for Judgment and Decision Making's mailing list. If unstructured interviews are not harmful, they could be useful if they convey some other benefit, such as increasing commitment from an applicant to accept if given an offer of admission. Second, even if interviewers do not consistently exploit it to increase overall accuracy, the unstructured interview may sometimes uncover important information that is special to a candidate. These unusual cues, akin to what Meehl (1954) famously called "broken legs," could immediately remove a candidate from consideration or catapult a candidate ahead of others. If an interviewer cares about some kinds of errors, such as missed broken legs, more than others, then unstructured interviews may be highly valuable. Basic psychological research, however, gives us reason to doubt that unstructured interviews will not harm judgment or that interviewers will be adept at spotting special information, and not false alarms, about a candidate.

Can interviews hurt? Access to an interview could hurt predictive accuracy because exposure to non-diagnostic information is known to *dilute* valuable information. Unstructured interviews expose interviewers to so many observations about the interviewee that have unknown or no diagnosticity that they receive literally more information than they can use. Research on the "dilution effect" (e.g., Nisbett, Zukier, and Lemley, 1981; Zukier, 1982; Peters and Rothbart, 2000) shows that rather than just being ignored, extraneous information reduces reliance on good information. It is perhaps no coincidence that the stimuli for the earliest dilution effect studies, which included ample material judged non-diagnostic by study participants, came from interview snippets (Nisbett, Zukier, and Lemley, 1981). Because making good social judgments often requires ignoring information and relying on simple rules, cognitive traits that might normally be construed as positive, like complexity of thought and need for cognition, can actually be detrimental to accuracy (Ruscio, 2000). A clever and capable interviewer may conduct an interesting unstructured interview, but be distracted from accuracy in doing so.

Can interviewers reliably extract special information from unstructured interviews? While too much irrelevant information dilutes the prediction process, it can also lead to unwarranted confidence due to *sensemaking*. People seek to impose order on events (Gilovich, 1991), so much so that they often see patterns in random sequences. As such, even the noisiest interview data are readily translated into a good story (see Dawes, 2001, Chapter 7) about the interviewee. Just as one can, post hoc, fit a "significant" statistical model to pure noise, interviewers have too many degrees of freedom to build a coherent story of interviewees' responses. If the interviewee gives a response that is

inconsistent with the interviewer's impression, the interviewer can dynamically reformulate that impression, perhaps asking follow up questions until hearing a set of responses that confirm an impression. Without structure, interviewers may not ask questions intended to disconfirm these impressions because people are inclined to "positive testing" – seeking information that confirms their hypotheses (Mynatt, Doherty, & Tweney, 1977; Bassok and Trope, 1984; Klayman and Ha 1987; Devine, Hirt, and Gehrke 1990; review in Sanbonmatsu, Posavac, Kardes, & Mantel, 1998).

Sensemaking could thus allow unstructured interviewers to feel they understand an interviewee almost regardless of the information they receive. Unfortunately, a feeling of understanding, while reassuring and confidence inspiring, is neither sufficient nor necessary for accuracy (Trout, 2002). Empirically, confidence and accuracy are often poorly related in interpersonal prediction contexts (Dunning, Griffin, Milojkovic, and Ross, 1990; Swann and Gill, 1997) and confidence has been shown to increase with information even in situations where accuracy does not (e.g., Andersson, Edman, and Ekman, 2005; Hall, Ariss, & Todorov, 2007). People may thus feel confident in the validity of unstructured interview impressions even if they are worthless.

We experimentally tested the roles of dilution and sensemaking in the context of using unstructured interviews to predict social outcomes. Study participants predicted the semester GPAs of other students based on biographical information and in some cases, an unstructured interview. In some conditions, the interviewee secretly used a random responding system to answer questions. Consistent with dilution, participants' GPA predictions were more accurate without the unstructured interview. Consistent with sensemaking, participants were just as subjectively confident in their interview

impressions when they conducted random interviews. Further, participants' predictions were worse than a single piece of information, prior GPA, that they had when making the prediction. We thus provide direct evidence that unstructured interviews are worse than invalid, they actually decrease accuracy.

Study 1

To explore whether interviews could dilute judgments and make them worse, we had student participants predict the semester grade point average (GPA) of two other students, one prediction with background information and an interview, the other with just background information. To explore whether interviewers sensemake, we developed a random responding system that the interviewees could use during the interview to see whether it would perturb predictive accuracy or subjective confidence in interview impressions.

Methods

Interviewers and interviewees

Interviewers were 76 undergraduate students at Carnegie Mellon University who were recruited through campus advertising and paid for their participation. We employed five Carnegie Mellon undergraduates (two female) as permanent interviewees. The interviewees ranged in age from 18 to 22, and represented multiple races, majors, and class standings. Two of the interviewees worked for two semesters, creating a total of 7 different semester GPAs to be predicted. Their prior cumulative GPAs and GPAs for the semester to be predicted are listed in Table 1.

Procedures

Participants were introduced to a randomly assigned interviewee and asked to conduct a 20 minute interview with the goal of predicting the interviewee's GPA for a given semester. An experimenter remained in the room during the interview to track time and answer any questions about the task. Prior to interviewing, participants were told the interviewee's age, major, class standing, and course schedule for the semester to be predicted. Participants were offered a break 10 minutes into the interview, during which they could formulate more questions to ask.

After the interview, the interviewee was excused and participants made their GPA predictions, which were to be kept confidential from the interviewee. Before making their predictions, participants were given the interviewee's cumulative GPA prior to the target semester and informed that prior GPA by itself was the best statistical model for predicting GPAs at this institution (Lewis-Rice, 1989). After the GPA prediction, participants answered a brief questionnaire (see Table 2) probing whether they got to know the interviewee and whether the interview provided useful information. Finally, 68 participants predicted the semester GPA for another target whom they did not interview using only the target's background information and prior GPA.

Interview conditions

The structure of the interview varied according to the participant's random assignment to one of three conditions. In the *accurate* condition (n = 25), participants could only ask closed-ended questions, i.e. "yes or no" or "this or that" questions. Interviewees answered these questions accurately. The *random* condition (n = 26) was similar except that after the midway break, the interviewee secretly responded on a

pseudo-random basis. Interviewees noted the first letter in the last two words of each question and classified them as category 1 (letters A through M) or category 2 (N through Z). If both letters belonged to the same category, the interviewee answered yes (or took the first option of a "this" or "that" question) and otherwise answered no. This system tends to equalize yes and no frequencies regardless of the proportion, p, of words sampled from category 1, because the probability of matching categories, $p^2 + (1 - p)^2$, is closer to .5 than p itself.

A lack of a significant difference in accuracy or survey answers between the accurate and random conditions might reflect the deficient quality of all closed ended interviews, in which case it would not imply that random interviews would be as good if one were allowed to formulate questions however one wanted. To account for this explanation, we also conducted a *natural* condition (n = 25) in which no closed-ended constraint was placed on the interview.

Study 1 Results

The validity (correlation with actual outcomes) of GPA predictions following interviews (r = .31) was significantly lower than the validity of prior GPA alone (r = .65; $t_{(73)} = 3.77$, $p < .05$, **d** = .43; Hotelling's method for dependent r with Williams' correction), information participants had when making their predictions. While worse than prior GPA, a validity of .31 compares favorably to that of unstructured employment interviews for predicting job performance (Campion, Palmer, & Campion, 1997). This comparison is not totally appropriate, e.g. because GPA could be more reliable than job

performance ratings, but it provides some evidence that our participants were not merely deficient interviewers.

Some of our interviewees were concerned that the random interview would break down and be revealed to be nonsense. No such problems occurred. Further, random responding did not perturb accuracy: Only the validity in the random condition (r = .42) was significantly different from zero, while validities in the accurate (r = .20) and natural (r = .29) conditions were not, though these 3 values did not differ significantly from each other. One concern is that random condition participants relied more on prior GPA because the interview was bad, thus inflating accuracy in the random condition because prior GPA was a strong predictor. This was not the case; GPA predictions were no more correlated with prior GPAs in the random condition (r = .54) than in the accurate (r = .53) or natural condition (r = .67).

Table 2 shows that the mean agreement with the statements "I am able to infer a lot about this person given the amount of time we spent together" (accurate = 2.72, random = 2.83, natural = 2.80) and "From the interview, I got information that was valuable in making a GPA prediction" (accurate = 3.00, random = 3.31, natural = 3.12) was similar across all conditions, with no significant differences emerging ($F_{(2, 73)}$ = .233 and 1.714, respectively). While comparisons of accuracy and subjective impressions yielded null results between random and truthful interviews, in both cases the direction was "wrong" – prediction accuracy and impressions of usefulness trended higher for random interviews.

Although participants judged the interview to be somewhat informative, predictions were actually less valid with interviews (r = .31) than without them (r = .61),

consistent with dilution. Because these correlations involved different judgments by the same participant, we tested the difference using hierarchical regression with participant random effects, regressing GPA predictions on obtained GPA, a dummy = 1 if an interview was conducted, the interview × obtained GPA interaction, and dummies representing interviewees. The interaction term was negative and significant (b = -.30, $t_{(60)}$ = -3.26, $p < .01$), meaning that predictions were indeed significantly less correlated with outcomes when an interview was performed.

Study 1 discussion

Consistent with sensemaking, a random interview did not perturb either GPA predictions or subjective impressions about the quality of the interview or the extent to which they got to know the interviewee. Consistent with dilution, participants made better predictions without an interview. While participants generally agreed that they got useful information from interviews, interviews significantly impaired accuracy.

Perhaps one reason that participants felt interviews were useful and made sense of them even when they were random is that they conducted them. If participants merely watched the interviews, rather than conducting them, would they be less prone to either or both effects? By having participants watch pre-recorded interviews, we could also directly assess whether they can tell random from accurate by informing of the possible that the interview they watched is random and asking which they saw.

Study 2

Rather than conducting the interview themselves, participants in study 2 watched a pre-recorded interview that another student had conducted. Because participants were

not allowed to ask their own questions, they may be less prone to confirming their own theories of the interviewee and thus less prone to sensemaking. If so, we might expect participants to be able to discern random from accurate interviews.

Methods

Participants and interviews

Participants were 64 undergraduate students at Carnegie Mellon University who were recruited through campus advertising and paid for their participation. Eight Carnegie Mellon undergraduates (five female) participated as interviewees and consented to having two interview sessions recorded (one random, one accurate) as stimuli for the study. Interviewees ranged in age from 19 to 21, and again represented multiple races, majors, and class standings. Table 1 lists their prior and obtained GPAs.

Procedures

Procedures were the same as in experiment 1, with the following exceptions. Prior to conducting the experimental sessions, we video-recorded 16 interviews (one accurate and one random for each interviewee, natural interviews were not use) conducted similarly to experiment 1, except that the random interview was now entirely random responding. Participants were randomly assigned to watch one of the 16 interviews via computer interface and predict the interviewee's GPA for a given semester. Each interview was randomly assigned to four different participants. The post interview question wording was amended slightly (see Table 3) to reference the interview that was watched and the Likert scale now ranged from 1 to 5 and included a "neither agree nor disagree" point. After the post-interview questionnaire, participants were informed that

their interview was randomly drawn from a pool containing half random interviews and asked to guess whether it was random or accurate.

Study 2 Results

GPA predictions were about equally correlated with actual GPAs as in study 1 (r = .28). Prior GPA, however, did not predict semester GPAs as well for the sample of interviewees used in study 2 (r = .37) and was not significantly more accurate than participant predictions. Though the procedure in study 2 is somewhat different, it is informative to consider that combining samples from both studies, prior GPA alone predicts significantly better than our participants do with interviews ($t_{(137)} = 2.59$, $p < .05$, **d** = .44).

Subjective impressions were again unperturbed by random responding, even though participants did not control the interview. Mean agreement with the statements "I am able to infer a lot about this person given the interview I just watched" (accurate = 3.47, random = 3.47) and "From watching the interview, I got information that was valuable in making a GPA prediction" (accurate = 3.66, random = 3.75) was again similar across conditions, with agreement in the random condition being equal or higher. As in study 1, GPA predictions relied on prior GPA about the same for random (r = .58) and accurate (r = .55) interviews.

Figure 1 tabulates participants' judgments of whether they saw an accurate or random interview across interview type. Participants correctly classified 66% of the interviews, significantly better than chance ($\chi^2_{(1)} = 8.33$, $p < .01$). This result, however, was largely driven by the participants judging all interviews to be accurate: accurate interviews were nearly always judged to be accurate (29/32), and more than half of

random interviews were judged accurate (19/32). The tendency to judge all interviews accurate was stronger than the tendency to be correct (McNemar's test, $\chi^2_{(1)} = 11.63$, $p < .001$).

Interestingly, we did not replicate findings of dilution in study 2, largely because no interview predictions, which were not handled differently in this study, were much less accurate ($r = .26$) than in study 1, while predictions following all interviews were about as accurate as in study 1 ($r = .28$). Of course, while interviews did not make predictions worse, they also did not make them significantly better. Combining the data from studies 1 and 2 and repeating the test of the interview/no interview difference from study 1, predictions were still more accurate overall without interviews than with them ($b = -.15$, $t_{(117)} = 2.32$, $p < .05$).

Study 2 discussion

Watching interviews did little to mitigate sensemaking; participants' predictive accuracy and subjective impressions were similar after watching random and accurate interviews, and they were more likely to see interviews as accurate whether they were or not. One objection to our interpretation of Studies 1 and 2 is the presence of experimental demand to use interviews. Because we took the trouble of having participants conduct or watch interviews for the majority of the study's duration, it is not unreasonable to assume that participants felt they should use the interview, regardless of their feelings about its validity. Of course, such demands may be present in real-world settings in which one is forced to conduct an interview for screening purposes. Still, one may wonder whether participants believed that interviews aided accuracy, a question we explore in study 3.

Study 3

Methods

One hundred sixty nine Carnegie Mellon University students completed this task as part of a larger study session. Participants were given descriptions of the methods and conditions used in Study 1 (except that the random condition was full random as in Study 2) and asked to rank the interview types (including no interview) in terms of how accurate a student's predictions would be following each.

Study 3 Results

The modal accuracy rankings were natural interview first, followed by accurate, random, and no interview, respectively, making the most accurate prediction type the least favored. This ranking was also the single most common, chosen by 57 (33%) of our participants. No participant ranked the natural condition last, while 56% of participants ranked no interview last. The dominance matrix in Figure 2 depicts how many participants ranked the interview type in the column over the type in the row. Even random interviews, which by definition contain misleading information, were preferred to no interview by 96 participants (57%). Thus, while interviews do not help predict one's GPA, and may be harmful, our participants believe that any interview is better than no interview, even in the presence of excellent biographical information like prior GPA.

Discussion

We set out to examine whether unstructured interviews could harm predictive accuracy and whether interviewers would believe they garnered useful information from the interview regardless of its quality. Consistent with dilution, study 1 showed that participants were better at predicting other students' GPAs when they were not given

access to an unstructured interview in addition to background information. Consistent with sensemaking, participants were unperturbed in their ability to make coherent impressions when the interviewee responded randomly, both in terms of the accuracy of their predictions and their confidence in their subjective impressions. Prior GPA alone was significantly more accurate than our participants, even though they had prior GPA at their disposal. Study 2 showed that even when watching rather than conducting an interview, participants were still somewhat prone to sensemaking. Finally, study 3 showed that participants believe that interviews will help in this context, so much so that they rate random interviews as being more helpful than no interview, which was in fact the best way to make predictions.

Our findings suggest a rethinking of the meaning of interview validity. The validity of predictions made by interviewers or by numerically incorporating interviews into a model is uninformative unless it can be directly compared to predictions made by the same people without an interview. On its face, the validity of our participants' predictions looks respectable, yet these same participants were able to predict better when they did not have an unstructured interview, and could have predicted better still if they just used prior GPA.

In addition to the vast evidence suggesting that unstructured interviews do not provide incremental validity, we provide direct evidence that they can harm accuracy. Because of dilution, this finding should be especially applicable when interviewers already have valid biographical information at their disposal and try to use the unstructured interview to augment it. Because of sensemaking, interviewers are likely to feel they are getting useful information from unstructured interviews, even when they are

useless. Because of both of these powerful cognitive biases, interviewers probably over-value unstructured interviews. Our simple recommendation for those who make screening decisions is not to use them.

References

Andersson, P., Edman, J., and Ekman, M. (2005). Predicting the world cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, 21, 565-576.

Bassok, M., & Trope, Y. (1984). Decision makers's strategies for testing hypotheses about another's personality: Confirmatory or diagnostic? *Social Cognition*, 2, 199-216.

Bloom, R.F., and Brundage, E.G. (1947). Predictions of success in elementary school for enlisted personnel, in Stuit, D.B. (ed.), *Personnel Research and Test Development in the Naval Bureau of Personnel*. Princeton: Princeton University Press, pp. 233-261.

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

Carroll, J.S., Wiener, R.L., Coates, D., Galegher, J., Alibrio, J.J. (1982). Evaluation, diagnosis, and prediction in parole decision making. *Law & Society Review*, 17(1), 199-228.

Dawes, R.M. (2001). Everyday irrationality**:** How pseudoscientists, lunatics, and the rest of us systematically fail to think rationally. Boulder, Colorado: Westview Press.

DeVaul, R., Jervey, F., Chappell, J., Caver, P., Short, B., & O'Keefe, S. (1987). Medical school performance of initially rejected students. *Journal of the American Medical Association*, 257, (1), 47-51.

Devine, P., Hirt, E., and Gehrke, E. (1990). Diagnostic and confirmation strategies in trait hypothesis testing. *Journal of Personality and Social Psychology*, 58, 952-963.

Doherty, M., Mynatt, C., Tweney, R., & Schiavo, M. (1979). Pseudodiagnosticity. *Acta Psychologica*, 75, 1-11.

Swann, W. B. & Gill, M.J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology*, 73, 747-757.

Gilovich, T. (1991). How we know what isn't so: The fallibility of human reason in everyday life. New York: Free Press.

Garfinkel, H. (1967). Common sense knowledge of social structures: The documentary method of interpretation in lay and professional fact finding. In H. Garfinkel (Ed.), *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Hall, C.C., Ariss, L., Todorov, A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Human Decision Processes*, 103, 277-290.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. Industrial and Organizational Psychology: Perspectives on Science and Practice, 1, 333-342.

Highhouse, S., Kostek, J.A. (in press). Holistic assessment for selection and placement. APA Handbook of Testing and Assessment in Psychology.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211-228.

Lewis-Rice, M. (1989). Marketing post-secondary educational programs with implications of higher education administration. Doctoral dissertation, School of Urban and Public Affairs, Carnegie Mellon University.

Meehl, P.E. (1954). Clinical versus statistical prediction; a theoretical analysis and review of the evidence. Minneapolis: University of Minnesota Press.

Milstein, R.M., Wilkinson, L., Burrow, G.N., and Kessen, W. (1981). Admission decisions and performance during medical school. *Journal of Medical Education*, 56, 77-82.

Nisbett, R., Zukier, H., and Lemley, R. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, 13, 248-277.

Ruscio, J. (2000). The role of complex thought in clinical prediction: Social accountability and the need for cognition. *Journal of Consulting and Clinical Psychology*, 68, 145-154.

Sanbonmatsu, D., Posavac, S., Kardes, F., and Mantel, S. (1998). Selective Hypothesis Testing, *Psychonomic Bulletin & Review*, 5, 197-220.

Trout, J.D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science*, 69, 212-233.

Table 1. Interviewees' prior and obtained GPAs.

	Interviewees							
Study 1	***1***	***2***	***3***	***4***	***5***	***6***	***7***	
Prior GPA	3.32	3.28	3.24	3.23	2.95	2.84	2.81	
Obtained GPA	3.80	3.08	3.71	3.34	2.68	2.69	3.35	
Study 2	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***
Prior GPA	3.69	3.38	3.29	3.29	3.23	3.05	2.83	2.65
Obtained GPA	3.84	3.80	4.00	2.83	2.65	3.59	3.00	3.31

Table 2. Study 1 post-prediction questionnaire. Mean agreement with statements on a 4-point Likert scale (1 = disagree, 4 = agree) with standard errors in parentheses.

	Accurate	Random	Natural
I am able to infer a lot about this person given the amount of time we spent together.	2.72 (.68)	2.83 (.47)	2.80 (.58)
From the interview, I got information that was valuable in making a GPA prediction.	3.00 (.65)	3.31 (.55)	3.12 (.60)

Table 3. Mean Likert responses (5 = strongly agree) to post experimental questions by condition with standard errors in parentheses.

	Accurate	Random
I am able to infer a lot about this person given the interview I just watched.	3.47 (.92)	3.47 (1.08)
From the interview I just watched, I got information that was valuable in making a GPA prediction.	3.66 (.94)	3.75 (.98)

Figure 1. Frequency of judged interview type by actual interview type.

	random	accurate	Total
guess random	13	3	16
guess accurate	19	29	48
Total	32	32	64

Figure 2. Dominance matrix in which cell frequencies are the number of participants who ranked the column method over the row method.

	Natural	**Accurate**	**Random**	**No interview**	***Total***
Natural	--	36	12	13	*61*
Accurate	128	--	28	22	*178*
Random	153	136	--	68	*357*
No interview	152	142	96	--	*390*
Total	*433*	*314*	*136*	*103*	*986*