



Faculty Publications

2009

A Brief Feedback Intervention for Diagnostic Overshadowing

David Wood

Brigham Young University - Provo, d_wood@byu.edu

Terence J.G. Tracey

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Medicine and Health Commons](#)

Original Publication Citation

Wood, D.S., & Tracey, T.J.G. (2009). A brief feedback intervention for diagnostic overshadowing. *Training and Education in Professional Psychology*, 3, 218-225.

BYU ScholarsArchive Citation

Wood, David and Tracey, Terence J.G., "A Brief Feedback Intervention for Diagnostic Overshadowing" (2009). *Faculty Publications*. 2928.

<https://scholarsarchive.byu.edu/facpub/2928>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

A Brief Feedback Intervention for Diagnostic Overshadowing

David S. Wood and Terence J. G. Tracey
Arizona State University

Clinical decision-making errors are well-documented among both experienced clinicians and students. One robust clinical decision-making error is called diagnostic overshadowing (DO), which occurs when the presence of one diagnosis interferes with the detection of other diagnoses. This study tested whether two types of instruction and brief feedback interventions reduced the likelihood of DO. Specifically, content-based feedback and principle-based feedback significantly reduced the likelihood of DO among doctoral students in clinical and counseling psychology ($N = 220$). An intervention effect was found when the training task and the target task were highly similar. Recommendations for improving diagnostic decision-making among trainees in professional psychology are discussed.

Keywords: clinical decision-making, psychodiagnosis, diagnostic overshadowing, feedback, training in professional psychology

Two daunting issues face mental health practitioners when they make clinical-decisions: 1) Clinical decision-making is an error prone process (Faust, 1986; Garb, 1998), and 2) Clinicians rarely get feedback on the accuracy and validity of their decisions so they have few opportunities to improve (Garb & Schramke, 1996). The prevalence of human information processing errors has been emphasized in the now classic work by Kahneman and Tversky (1974), who described several types of heuristics, or simplified and implicit judgment rules that are “quite useful but can lead to severe and systematic errors” (p. 1124). Focus on judgment bias and decision-making errors among mental health practitioners has also garnered attention. Many types of errors have been identified among mental health clinicians (Tracey & Rounds, 1999). Extensive reviews on clinical decision making, such as Garb (1998) and Faust (1986) have identified client characteristics related to biased decision making (race, social class, gender, sex role and age), as well as information processing errors (illusory correlations, context effects, labeling effects, overperception of psychopathology, overconfidence and diagnostic overshadowing). These reviews join with what has become a persistent critique of clinical decision making and the need to improve (e.g., Dawes, 1994; Lopez, 1989; Meehl, 1960). The focus of this study is on a specific type of

clinical error, that of overshadowing, which occurs when multiple disorders are present, but the salient features of one condition preclude or overshadow consideration of others. We focus on overshadowing and evaluate the efficacy of two brief feedback interventions in their ability to reduce overshadowing.

Diagnostic overshadowing—a term first coined by Reiss, Levitan and Szysko (1982) has been well established in analogue studies examining mental retardation (MR; Alford & Locke, 1984; Levitan & Reiss, 1983; Reiss et al., 1982; Reiss & Szysko, 1983; Spengler, Strohmer, & Prout, 1990). This literature shows that diagnostic overshadowing is a robust phenomenon among clinicians who, when presented with case studies involving MR, tend to overlook mood disturbance (Reiss et al., 1982), agoraphobia (Levitan & Reiss, 1983; Reiss et al., 1982), thought disorders (Garner, Strohmer, Langford, & Boas, 1994; Reiss et al., 1982; Reiss & Szysko, 1983; Spengler & Strohmer, 1994; Spengler et al., 1990), avoidant personality disorder (Reiss et al., 1982), and overall severity of psychopathology (Alford & Locke, 1984). White et al., (1995) conducted a meta-analysis of the MR overshadowing effect and found a small to moderate effect size of .19. The existence of diagnostic overshadowing for MR has been reported in clinical case studies as well (Hurley & Sovner, 1995; Tranebjaerg & Orum, 1991).

Diagnostic overshadowing has been examined in areas outside MR using analogue and field research methods. Analogue research designs have shown that traumatic brain injury and epilepsy overshadow psychiatric disorders (Garner et al., 1994), and that learning disabilities and hearing impairment overshadow behavior disorders (Goldsmith & Schloss, 1984). Field studies have shown a diagnostic overshadowing effect of mood and thought disorders on substance use disorders in psychiatric settings (Ananth, Vandewater, Kamal, & Brodsky, 1989; Goethe & Ahmadi, 1991; Skodol, Williams, Spitzer, Gibbon, & Kass, 1984). Conversely, several field studies conducted in substance use treatment settings have shown an overshadowing effect of substance use disorders on anxiety disorders (Fals-Stewart & Angarano, 1994; Kranzler et al., 1995; Riemann & McNally, 1992).

DAVID S. WOOD, PhD, holds a doctorate in Counseling Psychology from Arizona State University. He currently works as Research and Evaluation Specialist for LDS Family Services in Salt Lake City, UT. His area of focus in research is clinical decision-making and program evaluation.

TERENCE J. G. TRACEY received his PhD in Counseling from the University of Maryland, College Park in 1981. He has worked at the State University of New York at Buffalo and the University of Illinois at Urbana-Champaign, and currently is a Professor and Program Leader of the Counseling and Counseling Psychology programs at Arizona State University. His current scholarship focuses on vocational interest measurement and development and interpersonal models of personality and psychotherapy.

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to David S. Wood, 4722 W. Country Club Dr., Highland, UT 84003. E-mail: davidwood00@gmail.com

Several methods for reducing diagnostic overshadowing have been proposed. Generally speaking, Garb and Schramke (1996) have suggested that one reason why clinicians may make judgment errors is due to a lack of feedback. Feedback is assumed to have a powerful remedial effect on decision-making errors because it allows the clinician to learn from experience, as well as understand the impact of the clinician's decisions and actions (Lambert, Hansen, & Finch, 2001). We focused on the benefit of providing feedback in reducing errors. However, there are two separate pedagogical methods for providing feedback: content based and principle based.

Texts that focus on instruction of diagnostic skills (e.g., Fauman, 2002; First, Frances, & Pincus, 2002; Morrison, 2000) focus on symptoms, etiological rule-outs, and competing diagnosis rule-outs. Individuals are instructed on where they have made errors and then correct content is typically provided (e.g., exact criteria that apply to the diagnosis are provided). Generally, there is little focus on potential information processing errors in these texts. This more traditional approach to diagnosis was also represented in this study as *content-based* feedback. This type of feedback focused only on the specific symptoms that may have been overlooked by clinicians.

An alternative model that focuses more on learning to minimize information-processing errors was also examined. This feedback condition was labeled principle-based feedback because the information-processing error that was specifically used was of focus instead of the specific content of the diagnostic error. For example, a content-based feedback would focus on what diagnosis was overlooked in an overshadowing case and then highlight the criteria of this omitted diagnosis. A principle-based feedback would focus on the informing the participant of the overshadowing error itself and explain how this type of error is frequently committed. The principle-based feedback focuses on teaching the metacognition concept (Driscoll, 2000) rather than more case-specific content. We hypothesized that the principle-based feedback would result in less overshadowing especially in cases involving different diagnoses. There should be more generalization of learning in the principle-based feedback condition over a content-based condition given the broader focus on metacognition. In addition, both the content-based and principle-based feedback conditions should yield less overshadowing than a no-feedback control group.

So the focus was on evaluating a brief feedback intervention aimed at decreasing overshadowing in diagnosis. The brief intervention thus involved two separate feedback models: content- and principle-based. We hypothesized that both would result in less overshadowing relative to no-feedback control and further, that the principle-based feedback condition would demonstrate greater generalization of learning to other diagnostic types than would the content-based feedback condition.

Method

Sample

Graduate students in doctoral programs in clinical and counseling psychology were recruited as participants in this Internet-based study. A total of 220 participants completed the study. The mean age of the participants in the study was 29.49 ($SD = 5.59$).

Seventy-five percent of the sample was female. Regarding ethnicity, 76% of the sample was White, 6% Asian American, 5% multiethnic, 4% Hispanic, 4% African American, 2% international student, 1% Native American, and 0.5% other. Sixty-nine percent of participants held a masters degree, whereas 23% held only a bachelor's degree and 7% held a doctorate degree at the time of their participation. The majority of all respondents were currently enrolled as doctoral students (86%). Six percent indicated that they were enrolled in a masters program, 5% said they were in a postdoctoral program, and 2% indicated they were licensed professionals. Eighty-six percent of all participants said their clinical specialty was psychology, 9% indicated mental health counseling, 2% selected social work, 2% indicated marital and family therapy, and 1% said other. Given that only students in American Psychiatric Association (APA)-accredited programs were solicited to be in this study, the presence of these individuals with nonpsychology backgrounds was surprising and we interpreted this as reflecting the small number of individuals who indicated a nonpsychology specialty training prior to or in addition to their enrollment in an APA-approved program.

Stimulus Materials

All study materials were presented via the Internet. Participants signed on and were given a brief demographic questionnaire. They were then randomly assigned, based on their institutional program, to one of three conditions: principle-based feedback, content-based feedback, and no feedback control. The participants were then presented with two training vignettes, after each of which the feedback conditions were implemented. Following the two training vignettes, three outcome vignettes were presented. All vignettes were presented to all participants, with only the three feedback conditions (principle-based, content-based, or no-feedback) varying after the two training vignettes.

Participants were asked to read each vignette and make diagnostic decisions based on the information presented. The decision-making task was similar to that found in other diagnostic decision making studies, namely, participants made probability ratings of a list of diagnoses. For both the training and outcome vignettes in all conditions, participants made ratings of five diagnostic options according to the probability that the diagnosis was present in the case history (1 = not present; 4 = possibly present; 7 = definitely present).

Training Vignettes

Training vignette 1 ("Susan") was identical to that used in previous diagnostic overshadowing studies (Reiss et al., 1982) except the name and gender of the patient was changed to a female. Dr. Reiss was contacted and permission was granted to use the vignette for the study (Reiss, personal communication July 23, 2003). MR was the conspicuous diagnosis, whereas schizophrenia was the overshadowed diagnosis.

Training vignette 2 ("Ed") was contrived based on clinical reports described by Hurley and Sovner (1995), who indicated that the diagnosis of antisocial personality disorder is often overshadowed by MR. The fictional client had a history that clearly met the criteria for MR, which was the intended conspicuous diagnosis. He also displayed sufficient antisocial behavior, cognition and affect

to meet criteria for antisocial personality disorder, which was the intended overshadowed diagnosis.

Feedback Conditions

The intention of the *principle-based* feedback condition was to teach participants about a common information processing error (diagnostic overshadowing), as well as prompt a metacognitive strategy for the task (i.e., monitoring their own decision-making processes and cognitive strategies). Consequently, feedback was given in terms of whether or not the person made a diagnostic decision-making error, followed by a definition of diagnostic overshadowing. The statement for the first training vignette in the *principle-based* condition was:

The details in this vignette meet criteria for both MR and schizophrenia. If you rated schizophrenia as a 4 or less, you made a decision-making error called *diagnostic overshadowing*. Diagnostic overshadowing occurs when features of one diagnosis stand out more than features of another diagnosis (i.e., one overshadows the other).

For the *content-based* condition, the feedback made no mention of the term diagnostic overshadowing, but instead, listed the *DSM-IV-TR* (APA, 2000) criteria that were present in the vignette that also fulfill criteria for the overshadowed diagnosis. This can be construed as a “classic” approach to training in diagnostic decision making. The *content-based* feedback statement for the first training vignette was:

The details in this vignette meet criteria for both mental retardation and schizophrenia. If you rated schizophrenia as a 4 or less, you made a decision-making error. The details in this vignette meet the following *DSM-IV-TR* criteria for schizophrenia: A.1, A.3, A.4, and A.5.

Both the *principle-based* condition and the *content-based* condition were relatively balanced with respect to length. The *content-based* feedback condition would have been potentially much lengthier had each criterion been described in the text. Instead, the criterion number as it appears in *DSM-IV-TR* was provided. With this modification, the *principle-based* and *content-based* feedback statements were 51 and 45 words, respectively.

Lastly, the control condition included only information that the participant had made an error or not. In this condition, there was no elaboration as to why there was an error made. The feedback statement for the control condition is included below:

The details in this vignette meet criteria for both mental retardation and schizophrenia. If you rated schizophrenia as a 4 or less, you made a decision-making error.

For all conditions, a reference to a research study was provided that supported the feedback statement for the vignette. Participants were able to view an “answer key” to the most appropriate diagnostic ratings for each of the five options. This visual form of feedback highlighted the diagnoses that were not correct, as well as the two (the conspicuous and overshadowed) diagnoses that were correct. In addition to the text feedback and answer key, participants were able to click on a hyperlink and review the diagnostic ratings that they made for each training vignette.

It was anticipated that some participants would be more effective diagnosticians than others. To control for a ceiling effect, feedback was given for all options for each training case history, regardless of the correctness of the participant’s response. This

was intended to make the training effect more likely even for those individuals for whom diagnostic overshadowing was less of a concern. Additionally, the difficulty level of the training case histories was intentionally high, so that most individuals would have opportunity to see where they may have made overshadowing errors. Given the frequency of overshadowing among the numerous analogue and field studies, it was anticipated that most participants would make some errors.

Outcome Vignettes

The three outcome vignettes were written with (a) a conspicuous or overshadowing diagnosis, and (b) a less evident series of symptoms that meet criteria for the overshadowed diagnosis. Both the primary and the overshadowed diagnosis were listed as one of five options for which participants were asked to make probability ratings. The other three options were distractor items for which some, but not enough diagnostic criteria were met. The writing of the clinical vignettes entailed establishing the intended and non-intended diagnoses through carefully attending to *DSM-IV-TR* criteria and generating a written narrative that differentially included criteria for the diagnoses. After each diagnosis and its intended criteria were established, clinical vignettes from various books on diagnostic instruction (e.g., Fauman, 2002; Morrison, 2000) were used as examples to craft plausible details and the flow of clinical information for each vignette. Specifically, three different outcome vignettes were created: one to test overshadowing of the same content, (i.e., MR overshadowing another diagnosis), one to test generalization of learning to another example of diagnostic overshadowing besides MR, (i.e., specifically schizoid personality disorder), and a nonovershadowing conditions of a pure diagnosis. These three outcome vignettes were presented in the common order of: the generalization condition, (“Stephen,” schizoid personality disorder overshadowed by major depression, with the following as distractor diagnoses: delusional disorder, schizotypal personality disorder, and generalized anxiety disorder), the similar diagnostic overshadowing condition, (“Jason,” major depressive disorder overshadowed by MR with the following as distractor diagnoses: social phobia, attention deficit hyperactivity disorder, schizophrenia), and the pure diagnostic condition, (“Felicia,” alcohol abuse only with the following as distractor diagnoses: alcohol dependence, major depression, borderline personality disorder, and paranoid personality disorder). Following each vignette, respondents were requested to endorse the probability of diagnosis of five diagnoses using the seven point response format (1 = not present; 4 = possibly present; 7 = definitely present). Following the rating of the presence of all diagnoses, participants were asked to rate their confidence in the diagnostic ratings using a seven point scale (1 = no confidence, “I guessed”; 4 = somewhat confident; 7 = strong confidence - “I am confident my ratings are highly accurate”).

Validity Ratings

The three outcome vignettes were validated by a team of raters that included five advanced graduate students in counseling psychology and one tenured professor in counseling psychology. The graduate students were familiar with the *DSM-VI-TR* (APA, 2000) diagnostic criteria through graduate level course work and super-

vised practicum experience. The rating team was presented with the outcome vignettes as they appeared in the study. Raters were asked to determine whether or not each criterion was represented in the vignette. Rather than assign diagnoses, each rater had a list that included each behavioral diagnostic criterion as listed in the DSM for all five diagnoses that were listed as options for the outcome vignettes. These ratings were then taken and coded as to whether each rater's selections endorsed or failed to endorse minimum criteria for the diagnosis. The resulting ratings were compared with the rating criterion, which in this case was each intended diagnoses and each distractor diagnosis. As an example, for the "Stephen" vignette, the rating team was presented with the vignette as it appeared in the study. After reading through the vignette, the raters reviewed a list all the symptoms listed in the DSM associated with each of the five diagnostic options being assessed for this client and indicated whether or not the symptom was present in the vignette with a "yes" or "no." The list of symptoms was taken directly for each of the five diagnoses that would eventually be presented to the participants in the study. In this case, the symptoms were the diagnostic criteria for the following diagnoses: major depression, delusional disorder, schizoid personality disorder, schizotypal personality disorder, and generalized anxiety disorder. The actual diagnostic labels (e.g., major depression) were not provided to the rating team. These ratings were scored as to whether the responses resulted in a diagnosis or not for each set of symptoms on the rating form. For instance, if a rater indicated "yes" on five or more of the symptoms on the list that corresponded to major depression, the ratings were scored to indicate that the diagnosis was present. This was done for each of the criteria sets. Kappa statistics were computed for each rater with the criterion diagnoses. The mean kappa coefficient for each rater and the criterion ratings were then used as the final measure of representativeness of the diagnostic options for the vignette.

The "Stephen" vignette (schizoid personality disorder overshadowed by major depression) obtained an average pairwise kappa coefficient of .80 over the six raters, with a range of .55 – 1.00 and a *SD* of .23, which indicates a high level of agreement with the criterion diagnoses. The "Jason" vignette (major depressive disorder overshadowed by mental retardation) resulted in an average pairwise kappa coefficient of 1.0, which indicates perfect agreement with the criterion diagnoses. The "Felicia" vignette (alcohol abuse only) had an average pairwise kappa coefficient for each of the five diagnostic options of .70, with a range of .55 – 1.0 and a *SD* of .26, which indicates good agreement with the criterion diagnoses.

This approach of rating diagnostic criteria and not making diagnostic decisions is methodologically important because this reduces the probability of making inferential errors about diagnosis. Furthermore, Garb (1998) has argued that methods for validation of criterion cases in clinical-decision making studies should be methodologically distinct from the tasks required of participants. Otherwise, the results are merely reliability ratings between raters and research participants. Checklist-like structured clinical interviews have been demonstrated to be a method to avoid many types of clinical decision-making errors (Garb, 1998). Our usage of this behavioral checklist validation strategy provides more accurate information than having global ratings of diagnosis.

Procedures

Both clinical and counseling APA-accredited programs were solicited to participate in this study via electronic methods. Forty-one clinical psychology programs were randomly selected from a stratified, rank ordered list of 183 programs based on their cumulative scores on the Examination for the Professional Practice of Psychology (EPPP; Plous, S., n.d.). In order to assure an equal distribution of clinical programs from across all levels, this rank-ordered list was divided into quartiles, and random assignments were made to one of the three feedback conditions within each quartile. All known counseling psychology programs ($N = 83$) were requested to participate, and 17 either refused participation or could not be contacted electronically. The remaining 66 counseling psychology programs were randomly assigned to one of the three feedback conditions. While email verification was not required or requested from academic programs during this solicitation process, several programs responded and indicated that they would forward the solicitation to student listservs in their respective programs. Specifically, 19 counseling psychology programs and 10 clinical programs verified that they would send the solicitations out via an email message.

Each program was randomly assigned to a different feedback condition. The *principle-based* condition had 79 participants, the *content-based* condition had 96 participants and the control condition had 45 participants. Two-way contingency table analyses were conducted to determine if there was differential distribution of participants across feedback conditions. No significant relationships were found across feedback condition for gender, $\chi^2(2, N = 220) = 2.68, p = .26$, Cramer's $V = .11$; ethnicity (majority vs. minority), $\chi^2(2, N = 220) = 1.60, p = .45$, Cramer's $V = .09$; and degree earned (only BA vs. graduate), $\chi^2(2, N = 220) = 2.18, p = .34$, Cramer's $V = .10$.

Because participants were nested according to their clinical training program, there was a possibility of nonrandom distribution among the various categories. A test of the homogeneity of variance assumption was conducted to test whether or not this assumption was violated. The Levene statistic was used to test the null hypothesis that the error variance of the dependent variable is equal across groups. Type of feedback was the independent variable and included three levels: principle-based, content-based, and control. The dependent variable was the mean overshadowed diagnosis rating. The Levene test was not significant, $F(2, 217) = 2.76, p = .07$. Strictly interpreted, this test suggests that the population variances from which the samples were drawn were not significantly different and provides modest support that this assumption was not grossly violated in the inferential tests reported in this study.

Results

The means and standards deviations of the diagnostic ratings for each of the outcome vignettes are presented in Table 1. One of the analyses of interest involved the differences between the salient and the overshadowed diagnostic ratings. The scores for the salient diagnosis and the overshadowed diagnostic ratings were examined using a mixed-factor, repeated measures analysis of variance (ANOVA) with the five specific diagnostic ratings serving as the within subject variable (Salient diagnosis rating for Stephen, over-

Table 1
Participant Ratings of Dependent Variables According to Feedback Condition

Vignette/dependent variable	Feedback condition						Different groups
	Principle <i>M</i>	Based <i>SD</i>	Content <i>M</i>	Based <i>SD</i>	Control <i>M</i>	<i>SD</i>	
<i>N</i>	79		96		45		
MR/Axis-I ("Jason")							
Salient Diagnosis (MR)	6.10	0.90	6.35	0.73	6.13	0.92	CB > PB, CO
Overshadowed Diag. (Depression)	4.16	1.63	4.24	1.34	3.80	1.55	CB, PB > CO
Distractor Diagnoses	2.84	0.86	2.84	0.92	2.62	0.84	
Confidence Level	5.03	0.96	4.85	0.94	4.73	0.89	
Axis-I/Axis-II ("Stephen")							
Salient Diagnosis (Depression)	6.33	0.63	6.40	0.70	6.56	0.59	
Overshadowed Diag. (Schizoid)	3.96	1.78	4.22	1.47	3.69	1.88	CB > PB, CO
Distractor Diagnoses	2.26	0.79	2.33	0.83	2.26	0.98	
Confidence Level	5.10	0.96	5.00	0.93	4.96	1.11	
Diagnostically Simple ("Felicia")							
Salient Diagnosis (Alcohol Abuse)	5.86	1.37	6.32	0.88	6.09	1.43	CB > PB, CO
Distractor Diagnoses	3.14	0.63	3.10	0.61	3.18	0.69	
Confidence Level	5.32	0.84	5.30	0.85	5.29	0.97	

Note. All values are based on 1–7 rating scales. CB = Content-based Feedback; PB = Principle-based Feedback; CO = Control, no feedback.

shadowed diagnosis rating for Stephen, salient diagnosis rating for Jason, overshadowed diagnosis rating for Jason, and salient diagnostic rating for Felicia) and feedback condition (principle-based feedback, content-based feedback, and no-feedback control) serving as the between subject variable. The results of the ANOVA were significant for the within-subjects effect of differences across the five diagnostic ratings ($F(4, 214) = 192.03, p < .001, \eta^2 = .78$) indicating there were differences in the levels of diagnosis given across the ratings, which was hypothesized at least with respect to the overshadowed diagnoses. The post hoc tests indicated that the highest probability was found for the salient diagnosis of major depression for Stephen, which was significantly greater than the salient diagnoses of MR for Jason and alcohol abuse for Felicia, which in turn was significantly greater than the two overshadowed diagnoses, major depression for both Jason and schizoid personality disorder for Stephen. So across feedback condition, the participants saw the overshadowed diagnoses as significantly less probable. These results support the presence of overshadowing and add to the numerous studies already in the literature that document diagnostic overshadowing errors.

The between subjects feedback condition effect was also significant, $F(2, 217) = 3.92, p < .05, \eta^2 = .04$. Deviation contrasts were used to examine this effect because they are more interpretable given these data than the default polynomial contrasts because we specified contrasts between groups of interest that fit our questions, rather than using the default weighted composites of the groups. The deviation contrasts revealed that across all five diagnostic ratings, the content-based feedback group had the highest probability ratings and these were significantly greater than those of the other two groups. So across all salient variables, the content-based group was more likely to endorse the correct responses.

The primary question of the study involved the interaction between the feedback condition and diagnostic ratings. This interaction effect was significant, $F(8, 430) = 2.01, p < .05, \eta^2 = .04$. Post hoc Tukey's tests were conducted to determine the differences among the feedback groups for each of the five diagnostic

ratings. These post hoc tests are summarized in the right column of Table 1.

For the ratings of Jason, the vignette that was most similar to the training vignettes in that it involved MR as the salient diagnosis, the ratings of MR were significantly higher in the content-based feedback group than in either of the other two groups. This result makes sense as the specific content involved in the MR diagnoses was the feedback provided in this feedback condition. The overshadowed diagnosis of major depression was rated as significantly more probable in both feedback conditions than in the no-feedback control condition. The feedback of either type was related to less overshadowing.

The Stephen vignette was included as an examination of the generalization, as it involved a different salient diagnosis (depression). There were no significant differences on the salient diagnosis of major depression across feedback conditions. The overshadowed diagnosis of schizoid personality disorder was rated significantly higher in the content-based feedback condition than in the other two conditions. Content-based feedback was related to less overshadowing.

The final vignette was a "clean" description involving no overshadowing. The content-based feedback condition resulted in significantly higher probability ratings for the salient diagnosis than either of the other feedback conditions.

We conducted a similar repeated-measures ANOVA on the confidence ratings obtained for each of the three vignettes. There was a significant repeated-measures main effect for specific vignette ($F(2, 216) = 23.74, p < .001, \eta^2 = .18$). Post hoc tests indicated that the participants were significantly more confident of their ratings for the Stephen (major depression/schizoid personality disorder) than they were of their ratings of the Jason vignette (MR/major depression), and both of these ratings were significantly lower than the confidence ratings for the Felicia (alcohol abuse) vignette. There were no significant between subjects effects for feedback condition, $F(2, 217) = 0.67, p > .05, \eta^2 = .01$. So feedback condition was not related to confidence differences.

There was also no significant feedback by vignette interaction, $F(4, 434) = 0.72, p > .05, \eta^2 = .01$, indicating that the confidence ratings did not vary across the feedback conditions differentially by vignette.

There was a possibility that the different feedback conditions had the undesirable effect of alerting respondents to make diagnoses that were not present in each case. The design of this study included distractor diagnoses, which involved diagnoses representing some of the diagnostic criteria, but not sufficient to meet minimum criteria. For each of the vignettes, the distractor diagnosis ratings were tested across the different conditions. In each instance, none was significantly different from each other, indicating that the feedback did not have this undesirable effect: Jason, $F(2, 217) = 0.16, p > .05, \eta^2 = .01$; Stephen, $F(2, 217) = 1.10, p > .34, \eta^2 = .01$; Felica, $F(2, 217) = 0.26, p > .05, \eta^2 = .01$.

Discussion

Consistent with results from the body of literature on this topic, diagnostic overshadowing was apparent in this study (e.g., Reiss et al., 1982; Spengler et al., 1990). The brief feedback intervention presented here, however, showed a significant effect, suggesting that the tendency to make the diagnostic overshadowing error can be minimized through training and feedback. Specifically, diagnostic overshadowing is reduced by either *principle-based* feedback or *content-based* feedback, but only when the training task is highly similar to the target task. This study found no appreciable differences between the *content-based* and *principle-based* models of decision making, but these two conditions resulted in significantly less overshadowing than the control group. Also, this study found that a generalization effect with the *content-based* feedback occurred. Another important implication of this study is that a brief feedback intervention for diagnostic overshadowing can be delivered in an electronic format, which can either stand alone or supplement other types of training.

Metacognition was used as a conceptual framework to explain why *principle-based* feedback should be superior to that of *content-based* feedback. It was hypothesized that metacognition, or the ability to self-monitor and select appropriate strategies, is the primary skill that clinicians must improve if they are to reduce the likelihood of making diagnostic decision-making errors. With respect to diagnostic overshadowing, however, this prediction was not borne out. In contrast, the *content-based* condition seemed to draw upon another important conceptual framework for diagnostic overshadowing: cognitive complexity. Cognitive complexity, or the ability to simultaneously process multiple constructs and their interrelationships (Bieri et al., 1966), has been demonstrated as a distinguishing variable for those who do and do not show proneness to make overshadowing errors (Spengler & Strohmer, 1994). The *content-based* feedback condition may have primed participants to attend to the diagnostic criteria more carefully, which is a rudimentary but often overlooked remedy for diagnostic errors (Garb, 1996). Diagnostic overshadowing was conspicuously absent among the rating teams who developed the master ratings for the posttest vignettes, suggesting that a criterion by criterion review of a wide range of possible diagnoses may also reduce diagnostic overshadowing. Lastly, the content-based feedback ap-

proach also showed a generalization effect, indicating that if training is focused on highly specific diagnostic criteria, individuals tend to be less likely to make the diagnostic overshadowing error when working with diagnoses other than those used in training. Furthermore, the *content-based* feedback approach may be more suited to trainees who are becoming familiar with DSM diagnostic criteria. The utility of *principle-based* feedback might not be fully realized until a diagnostician obtains substantial experience with the DSM.

Construct validity must be addressed in interpreting these results. Both the brevity and the format of the intervention may have contributed to the results of this study. Although brevity was prioritized in order to maximize participation and minimize fatigue, it is likely that increasing the length of the intervention might more adequately provide participants the opportunity to thoroughly learn and implement corrective strategies to avert diagnostic overshadowing. With respect to external validity, some research has shown that the probability for errors is equal in both graduate students and practicing professionals regardless of experience (Garb, 1989). Additionally, a sample size of 220 participants is not small on the face of it, but when considered in the context of all clinical and counseling trainees in over 260 programs, it is probably less than 5% of all possible participants.

One limitation of this study is the lack of true random assignment. Although all programs were randomly assigned to one of the three feedback conditions, participants that belong to the same program were assigned to (and nested in) the same condition. The assignment of programs to condition, instead of assigning individuals, was done to avoid the possibility of intersubject communication. Similarly, the name of each individual's training institution was not requested due to the possibility of being identified from among the sometimes small number of clinically trained students at some programs. One significant limitation of this study that complicates causal inferences is the omission of each participant's program of study because relative differences in the quality of clinical training, if any, were not controlled. On balance, programs were rank ordered into quartiles and randomly assigned to conditions according to their cumulative averages on the EPPP. The EPPP was regarded in this study as a measure of general knowledge in psychology (the EPPP is a broad exam that has a subsection on psychodiagnostic knowledge). Because the relative differences in EPPP scores among the various programs were accounted for prior to random assignment, a modest albeit imperfect effort was made to protect against nonrandom distribution of participants to the various conditions. Future research is needed to see if similar results hold and, more particularly, to test for similar effects while controlling for nesting of participants within academic programs. Additionally, this study examined only the short-term effects of training and feedback. Future research should examine if training effects hold over time.

Noteworthy criticisms of online research have been made (Gosling, Vazire, Srivastava, & John, 2004). Among the many concerns noted is the validity of participants and participant responses. In other words, there are possible limitations due to not ensuring that respondents are actually who they say they are. Because of the confidentiality required by this study, as well as the extensive time requirement involved in validating participant identity, the study relied on sending solicitations directly to directors of training. Although this is not a complete guarantee, it offered some degree

of assurance that the solicitations were only received by actual graduate students in clinical fields.

The results of this study have specific implications for training in professional psychology, particularly in the area of mental health diagnosis. Training in psychodiagnostic skills should include 1) the disciplined practice of evaluating a wide range of diagnostic criteria each time a diagnosis is being considered, and 2) vigilance and critical self-evaluation for high likelihood decision making errors, including diagnostic overshadowing. In short, trainees will benefit from specific feedback and training on the possible pitfalls of heuristics. Specific feedback is likely best delivered during the actual diagnostic task. It may not be sufficient to merely read about decision-making errors. Based on our results, trainers can take a measurable degree of confidence that utilization of specific feedback in training can reduce the possibility of diagnostic overshadowing errors.

Clinical decision-making research plays a critical role in psychological assessment literature. It may be easy to conclude that information processing errors are inevitable and perhaps overly intractable, but the current study and hopefully those to come will contribute to a library of feedback interventions that successfully improve clinical decision making and enable clinicians to improve, monitor, and provide more suitable mental health diagnosis and treatment.

References

- Alford, J. D., & Locke, B. J. (1984). Clinical responses to psychopathology of mentally retarded persons. *American Journal of Mental Deficiency, 89*, 195–197.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- Ananth, J., Vandewater, S., Kamal, M., & Brodsky, A. (1989). Missed diagnosis of substance abuse in psychiatric patients. *Hospital and Community Psychiatry, 40*, 297–299.
- Bieri, J., Atkins, A. L., Briar, S., Leaman, R. L., Miller, H., & Tripodi, T. (1966). *Clinical and social judgment: The discrimination of behavioral information*. New York: Wiley.
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.
- Driscoll, M. P. (2000). *Psychology of learning for instruction* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Fals-Stewart, W., & Angarano, K. (1994). Obsessive-compulsive disorder among patients entering substance abuse treatment: Prevalence and accuracy of diagnosis. *Journal of Nervous and Mental Disease, 182*, 715–719.
- Fauman, M. A. (2002). *Study guide to DSM-IV-TR*. Washington DC: American Psychiatric Association.
- Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17*, 420–430.
- First, M. B., Frances, A., & Pincus, H. A. (2002). *DSM-IV-TR: Handbook of differential diagnosis*. Washington DC: American Psychiatric Association.
- Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387–396.
- Garb, H. N. (1996). The representativeness and past-behavior heuristics in clinical judgment. *Professional Psychology: Research and Practice, 27*, 272–277.
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Garb, H. N., & Schramke, C. J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analyses. *Psychological Bulletin, 120*, 140–153.
- Garner, W. A., Strohmer, D. C., Langford, C. A., & Boas, G. J. (1994). Diagnostic and treatment overshadowing bias across disabilities: Are rehabilitation professionals immune? *Journal of Applied Rehabilitation Counseling, 25*, 33–37.
- Goethe, J. W., & Ahmadi, K. S. (1991). Comparison of diagnostic interview schedule to psychiatrist diagnoses of alcohol use disorder in psychiatric inpatients. *American Journal of Drug and Alcohol Abuse, 17*, 61–69.
- Goldsmith, L., & Schloss, P. J. (1984). Diagnostic overshadowing among learning disabled and hearing impaired learners with an apparent secondary diagnosis of behavior disorders. *International Journal of Partial Hospitalization, 2*, 209–217.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*, 93–104.
- Hurley, A. D., & Sovner, R. (1995). Six cases of patients with mental retardation who have antisocial personality disorder. *Psychiatric Services, 8*, 828–831.
- Kahneman, D., & Tversky, A. (1974). On the psychology of prediction. *Psychological Review, 80*, 237–251.
- Kranzler, H. R., Kadden, R. M., Bureson, J. A., Babor, T. F., Apter, A., & Rounsaville, B. J. (1995). Validity of psychiatric diagnoses in patients with substance use disorders: Is the interview more important than the interviewer? *Comprehensive Psychiatry, 36*, 278–288.
- Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69*, 159–172.
- Levitan, G. W., & Reiss, S. (1983). Generality of diagnostic overshadowing across disciplines. *Applied Research in Mental Retardation, 4*, 59–64.
- Lopez, S. R. (1989). Patient variable biases in clinical judgment: Conceptual overview and methodological considerations. *Psychological Bulletin, 106*, 184–203.
- Meehl, P. E. (1960). The cognitive activity of the clinician. *American Psychologist, 15*, 19–27.
- Morrison, J. (2000). *DSM-IV made easy*. New York: Guilford Press.
- Plous, S. (n.d.). Ranking of clinical psychology programs in the U.S. and Canada. Retrieved December 10, 2003, from <http://www.socialpsychology.org/clinrank.htm>
- Riemann, B. C., McNally, R. J., & Cox, W. M. (1992). The comorbidity of obsessive-compulsive disorder and alcoholism. *Journal of Anxiety Disorders, 6*, 105–110.
- Reiss, S., Levitan, G. W., & Szysko, J. (1982). Emotional disturbance and mental retardation: Diagnostic overshadowing. *American Journal on Mental Deficiency, 86*, 567–574.
- Reiss, S., & Szysko, J. (1983). Diagnostic overshadowing and professional experience with mentally retarded persons. *American Journal on Mental Deficiency, 87*, 396–402.
- Skodol, A. E., Williams, J. B. W., Spitzer, R. L., Gibbon, M., & Kass, F. (1984). Identifying common errors in the use of *DSM-III* through diagnostic supervision. *Hospital and Community Psychiatry, 35*, 251–255.
- Spengler, P. M., & Strohmer, D. C. (1994). Clinical judgmental biases: The moderating roles of counselor cognitive complexity and counselor client preferences. *Journal of Counseling Psychology, 41*, 1–10.
- Spengler, P. M., Strohmer, D. C., & Prout, H. T. (1990). Testing the robustness of the overshadowing bias. *American Journal on Mental Retardation, 95*, 204–214.
- Tracey, T. J. G., & Rounds, J. (1999). Inference and attribution errors in test interpretation. In R. K. Goodyear & J. W. Lichtenberg (Eds.), *Test*

- interpretation: Integrating science and practice* (pp. 113–131). Boston: Allyn & Bacon.
- Tranebjærg, L., & Orum, A. (1991). Major depressive disorder as a prominent but underestimated feature of fragile X syndrome. *Comprehensive Psychiatry*, 32, 83–87.
- White, M. J., Nichols, C. N., Cook, R. S., Spengler, P. M., Walker, B. S., & Look, K. K. (1995). Diagnostic overshadowing and mental retardation: A meta-analysis. *American Journal on Mental Retardation*, 100, 293–298.

Received January 7, 2008

Revision received April 20, 2009

Accepted May 6, 2009 ■

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.