DOCUMENT RESUME

ED 358 465 CS 213 871

AUTHOR	Campbell, Elizabeth Humphreys
TITLE	Fifteen Raters Rating: An Analysis of Selected Conversation during a Placement Rating Session.
PUB DATE	Mar 93
NOTE	19p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (44th, San Diego, CA, March 31-April 3, 1993).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Freshmen; Ethnography; *Evaluation Methods; Evaluation Problems; Freshman Composition; Higher Education; *Holistic Evaluation; Scoring; Student Placement; *Writing Evaluation; Writing Research
IDENTIFIERS	Placement Tests

ABSTRACT

An ethnographic study investigated conversations during planning meetings and placement rating sessions for selecting and rating freshman composition placement exams. Planning meetings involved the administrative coordinator and two assistants selecting model essays, discussing the rubric, and confirming the final plans for the rating session. Fifteen instructors of first-year composition at a large, public, urban midwestern university holistically rated about 2,000 placement essays over a 4-day period. Conversations were tape recorded, transcribed, and analyzed. Results indicated that, while choosing the model essays, the planning team focused on an informal rubric; during the rating session itself, however, no mention of the informal rubric was made. Most discussions took place around problematic or puzzling essays; ones that fit clearly into a category did not require much discussion. Findings suggest that administrators of holistic assessment sessions should expect discussions during rating sessions to stray far from the terms of the written rubric as the raters struggle to work out the meaning of scores they assign. (A figure listing terms from the "mental rubric" and a figure presenting the written rubric are included.) (RS)

Fifteen Raters Rating: An Analysis of Selected Conversation During a Placement Rating Session

Elizabeth Humphreys Campbell

The University of Tennessee at Chattanooga

Abstract

Holistic assessment has previously been an opaque process; most research has looked at the results rather than the process. The conversation that took place during the planning session for a placement rating session and the conversation during the rating session itself were tape-recorded, transcribed, and analyzed. While choosing the model essays, the planning team focused on an informal rubric; during the session itself, however, no mention of the informal rubric was made. Most discussions took place around problematic or puzzling essays; ones that fit clearly into a category did not require much discussion. This discussion sheds light on the written and unwritten rules of holistic assessment.

This ethnographic study took place over a ten-day period in August, 1990, at a large, public, urban, midwestern university. Fifteen instructors of first year composition rated about 2,000 placement essays. Although the primary focus of the study is on the placement exam rating session, which took place over a four-day period, the background is drawn from observations of the planning meetings during the week before the session when the administrative coordinator and his two assistants met to select model essays, discuss the rubric, and confirm the final plans for the session.

The planning meetings and the placement rating sessions were tape recorded, transcribed, and analyzed. The conclusions of this study are based on those transcripts as well as questionnaires and interviews with members of the placement rating team.

At the time of the study, the English Department at "Midwest University" consisted of 48 full-time, tenure-track faculty; 4 adjunct assistant professors; 2 instructors; 1 visiting instructor; 22 graduate teaching assistants; 32 adjunct instructors; and 12 student lecturers (most of whom were A.B.D.). Most of those outside the tenure track taught first year composition courses almost exclusively, although a few taught sophomore-level composition or literature survey courses. There were also 4 graduate fellows in non-

teaching positions, 120 graduate students, and about 325 undergraduate majors.

Holistic assessment of placement exam essays had been the standard practice of the department I studied for about ten years; incoming students were placed into advanced, regular, or developmental composition sections, based on the results of the placement rating.

The session was organized and led by an administrative coordinator and two assistants. They were adjunct instructors who had been teaching in the department for several years and were very experienced in holistic assessment.

The Planning Meetings

The planning meetings took place during the week before the session; the administrative coordinator and his two assistants met to select 20 model or anchor essays, discuss the rubric, and confirm the final plans for the session.

The most striking fact about the planning meetings was that, in order to sort, categorize, and rank the essays, the team members used a sort of "mental rubric" to help them locate models with particular characteristics. This mental rubric was informal in the sense that it was not written,

but it was clearly an important part of the team members' mental constructs and guided their judgments during the selection process. Like an evolving mental scavenger hunt, the search for particular essays was guided by a list that was articulated as the session progressed. This group had worked together in previous years and found that selecting model essays that met specific criteria was an effective way to get their ideas across to the raters. The mental rubric was clearly part of their cultural knowledge even though it was unwritten. One member of the team said:

> We did that last year and I thought it worked out really well, too. We had these low 2's for different reasons. The minimum level in sentence errors, the minimum level in cliches and platitudes, the minimum level in organization or something... (Fieldnotes, p. 46 [FN])

In one case, the leader was describing a particular essay that he was looking for: "Well...what we need here is a rambling 2, kind of...vague, uneven, superficial... generalities" (FN, p. 61). After selecting the top and bottom models, these types of mental categories were consistently used by the group to determine which essays would fill in the middle ranges.

Figure 1: TERMS FROM THE MENTAL RUBRIC

Essays are rated holistically on a four-point scale

One Essays (Developmental English):

*the real low end 1's.
*totally superficial, not superficial in a two-ish way.
*as a high 1? There aren't really sentence level problems per se.
*really interesting as a 2/1 because if I'd seen a lot in the stack and you hadn't and I gave it a 1, it would go to a third reader.

Two Essays (Regular Composition):

*This certainly does ramble. It might be a very good example.
*organizationally, this essay is very weak...the minimum we'll take as far as organization.
*a 2 with organizational problems.
*this is the minimum we'll take as far as errors.
*...bad spelling problems. It might be interesting to use as a 2 with errors.
*the anecdote is the tail that wags the dog. Which is a wonderful 2 quality.
*a low two-ish, everyman sort of ring to it.
*a good, solid, typical, boring, 2. That thousands of others are going to be like.
*you know, one that you're going to get for the very first paper. No specificity...
*one that shows you can pass and not be that specific.
*For the 2/1? A 2 that never gets specific, relies on generality.
*what we need here is a rambling 2, kind of a vague, uneven, superficial, generalities.
* Maybe we need one that...meanders?
*kind of a shortish essay that passes.
*it's long enough...the student writer is able to sustain something and is able to discuss it.
*This is a wonderful 2/3 split...The last page just collapses.
*2 or 3 really glaring errors [are] consistent enough to keep 'em out of advanced...

Three/four Essays (Advanced Composition):

*"Competently written"? ...I mean that as a higher competent, not sparkling.
*It has that nice voice to it.
*the structure is the thing that tells us this student ought to be in advanced.
*much more conventionally academic. I mean not certainly in a bad way.

Figure 2: THE WRITTEN RUBRIC
[Distributed to raters at the beginning of the rating session]

4. the "4" essay is most often characterized by a sophisticated control of the elements of an essay writing situation. the essay addresses the topic and has a strong, sometimes subtle structure. the relationship between sentences and paragraphs results in a complex response. the ideas are well developed by specific details and concrete examples. it is generally free of mechanical errors.

3. the "3" essay is characterized by an effective control of the elements of an essay writing situation. it usually addresses the topic and is clearly structured. the relationship between sentences and paragraphs results in a well developed response. the ideas are usually developed with specific details and examples. it may contain random or sporadic mechanical errors, but they are not of sufficient severity or frequency to interfere with the expression of the ideas.

2. the "2" essay is often characterized by an uneven control over the elements of an essay writing situation. while it may not consistently address the topic, there is a sense of essay structure. the relationship between sentences and paragraphs may result in a superficial response. its ideas are usually developed by generalizations rather than specific details and examples. words are generally used accurately, although the essay may contain minor lapses in standard written english: spelling, punctuation, grammar or sentence structure.

1. the "1" essay is characterized by a lack of control over the elements of an essay writing situation. although the topic may be addressed, essay structure is usually weak (or absent). the relationship between sentences and paragraphs often results in an incoherent or incomplete response. its ideas are underdeveloped, fragmented, or stated as cliches or platitudes. most importantly, this essay usually contains serious or systematic errors in punctuation, grammar, spelling conventions, and/or sentence structure. it may be unacceptably brief.

The mental rubric provided the group with frames of reference in their search for several particular types of

essays. For example, in looking for one essay filled with cliches and platitudes, the leader found one that read, in part:

> Life is not easy, but it is what one makes it...The trick-- to be able to resist drugs and other things that corrupt your mind is simple-- just say no. Also believe in yourself and others will believe in and respect in your choice just to say no. One must also have faith to wait for the good things in life instead of wanting it all now...good things come to those who wait...there are two roads from which one must choose only one. (FN, p. 130)

Although it is clear that the essay quoted is filled with cliches and platitudes, one of the team members responded that she "would feel real uneasy placing this student in developmental" (FN, p. 45). Although an essay may fit into a category of the mental rubric for one reason, it may have other qualities that place it into a different category, and the team has to agree which factors are the determining ones.

The administrative team also had a strong desire to present a unified front. They considered even minor disagreements among themselves as portents of possible upheaval during the rating session. Part of the process of

assuring consensus in the large group required them to eliminate any models that did not fit their mental rubric. During the placement rating session, the mental rubric was never mentioned; an attempt was made to make sure that the language of the written rubric was consistently used during the rating session, but terms from the mental rubric did occasionally slip into the discussion.

The Rating Session

The rating session itself took place the following week. Fifteen raters participated in the 1990 placement rating session, six males and nine females.

Two of the three members of the administrative coordinating team had participated in the annual placement rating sessions for ten years. When added to other large-scale holistic assessments, by their own estimations, they had each participated in from 25 to 38 rating sessions lasting from a half-day to several days.

All other participants were also highly experienced at holistic rating. They had participated in at least 3 and up to 28 holistic assessment sessions.

There were nine adjunct or visiting instructors who participated in the rating session. Six hold the M.A. and

three the Ph.D. Of the six doctoral student raters, two were doctoral candidates at the time of the rating session and four were still engaged in doctoral level classes; all were either graduate teaching assistants or graduate fellows.

The raters listed several motivations for participating in the placement rating session: "to help determine to some degree the makeup of courses I teach"; getting "a realistic picture of incoming [students] and their abilities"; "hearing other teachers' views of the ideal student writer"; "the inadvertent student humor in the writing"; "enjoy wit of colleagues"; "staying current with expectations of student writing and pedagogical theories and practices"; "practically the only chance to share views with colleagues about the goals/evaluation of essays"; "it represents the overall impression of a hyper-aware reader"; "going with my first impression of a piece of writing and having those judgments corroborated by other raters"; "the chance to get to know what's happening with my friends"; "the camaraderie that comes out of agreement"; "a consensus is established."

One common reason for participating in "placement" is to get "a realistic picture of incoming [students] and their abilities." The administrative coordinator said:

> When you're confronted with students on the first day of class, you're going to think about what you

> saw in placement, and how those essays are connected. You know where the differences lie. If you're teaching developmental or advanced, you can see how to plan and teach the class much more clearly, based on having seen the whole sample.

One of the most interesting findings of this study was a distinct difference in the conversational patterns among the episodes. As one would expect, the shortest episodes had the highest levels of agreement; in most cases, when all the raters agreed upon a rating, they only needed to affirm their reasons for doing so. Conversely, when there was disagreement, the episodes were lengthy and sometimes impassioned. The *topics* of these conversations, however, proved to be very interesting.

A consistent pattern emerged: there is a striking difference between the way the raters talk about essays when they agree and when they disagree. When they agree, they talk about the model essay itself; they discuss the structure, the style, the meaning, the theme, or some other "objective" element. When they disagree, it is usually because the essay is a "problematic" one, and they begin more extensive narrations, making claims of professional expertise, relating stories from their teaching experiences and from past rating sessions; they construct the reader,

discuss the rubric, and discuss assessment theory in general.

As long as the raters are in agreement, the discussions center around "objective" criteria like sentence structure and word choices, but when the problems with the essay are not easily definable, they switch to "subjective," experience-based criteria. Of course, knowledge of such topics as essay structure and style are also based in experience, but the nature of the discussions is distinctly different.

The following quotations are taken from discussions during which the raters agreed on the rating of an essay and concerned themselves with such topics as word usage, sentence structure, coherence, types of comparisons, transitions, and "mechanics." Furthermore, there was no attempt to guess the writer's state of mind, speculate about the writer's personality, or examine her motivation.

> One rater commented: I think he has better control of mechanics and the sentences are far better, although there are some problems here or there. One of them might be that he tends to lean towards jargon. (FN, p. 104)
>
> Another rater commented: It's filled with cliches and platitudes. (FN, p. 110)

> Another commented:...there aren't transitions and it is a major flaw in here...those paragraphs are not explicitly connected. They are implicitly connected...It made me give it a 2 instead of a 3. (FN, p. 115)
>
> Another commented: It stays together. She sticks to her point. (FN, p. 125).

In contrast, when confronted with a puzzling essay, raters often attempted to provide possible explanations for writers' lapses.

> One rater said: I think what's attractive about this is that she does seem to care about what she's writing about -- especially toward the end, and it makes you want to sort of overlook a lot of other things, but I think if you rely on the rubric, it's really closer to a 2. (FN, p. 101)

In several cases, the raters expressed the desire to talk to the writers and find out what they were trying to accomplish:

> One rater said: I just thought...it did do some sophisticated things, but it existed too much on the level of generalization for me. And I kept wanting to say, you know, so give me an example - tell me what you mean by this.

> The coordinator replied: One real name! [heavy emphasis, general laughter] (FN, p. 105)

In some cases, the raters tried to imagine the writer's personality and figure out motives:

> One rater said: He's a brown-noser. I mean [he has] a very strong sense of what he's supposed to do in an educational situation. (FN, p. 106)
> Another said: I get the sense...that he's trying to impress somebody. He's done some reading. He's obviously knowledgeable in some areas. And he's trying so hard to impress the reader that he just got all screwed up in his sentence structure, and linking his sentences and thoughts together. (FN, p. 280)

In some cases, there was speculation about the student's physical state:

> The coordinator said: I think...the hard part of the essay is that it takes it a while to get going. These students mostly took these exams at 8:00 - in the morning - were given the topic, and in 45 minutes were expected to produce a piece of writing. I think that's a pretty difficult task that we gave them. A lot of these essays will take a while to get going. I'd like to caution you not to make up your mind about an essay too quickly -

> to make sure and read the essay all the way through. (FN, p. 101)
>
> Another rater said: Somebody getting out of bed in the morning is not pushing the absolute limit. (FN, p. 292)
>
> Another rater said: [the student]- drank heavily the night before this exam, or the morning of it. (FN, p. 292).
>
> Another commented: He probably talked to his brother who took it before. (FN, p. 203)

Another rater speculated about a student's general knowledge, basing the speculations on sketchy evidence:

> Certainly writing a five paragraph essay is not enough reason to fail it. I do think the paragraphs are underdeveloped, I think that the sentence structure is incredibly simplistic, on the whole, and it takes the same form throughout. He's - from what I can tell from this, he only knows how to write one paragraph, and that not terribly well, based on the supporting paragraphs he has in the middle of his paper. (FN, p. 220)

Finally, the attempt to empathize with a student writer is perhaps best illustrated by the following comment that obviously comes from many years of experience in testing situations:

> 41: Just throw out the last page, I mean, the guy said it's time to hand in your paper. Get rid of the last two paragraphs. (FN, p. 292)

In conclusion, these conversational patterns reveal that, for these raters, the routine essays do not require much discussion. Good writing or bad writing, when it is obvious, brings the raters into immediate consensus. This underscores the importance of selecting some model essays that clearly fit into categories described by the rubric; the raters need clear models to illustrate the idealized descriptions in the rubric. Yet, the majority of essays selected as models during this study fit the "problematic" category, and if this rating session is typical, administrators of holistic assessment sessions should expect those discussions to stray far from the terms of the written rubric as the raters struggle to work out the meaning of scores they assign; their teaching experience, professional expertise, "department standards," assessment theory, their ability to construct the writer - all these factors are part of the interactive context of rating placement essays.

While the raters are clearly aware of the traditional elements of style found in their composition handbooks and textbooks, they have also have acquired extensive "local knowledge." In their many years of teaching writing they have learned to "fill in the blanks" left by beginning

writers, to speculate about what is not easily visible, to wonder about the writers' motivations, and above all, to give them the benefit of the doubt. What rings through the transcripts again and again is that these instructors cared very deeply about the process they were engaged in; they saw the writers as real people with talents, ambitions, limitations, interests, prejudices, blind spots, and wisdom. And above all, they saw themselves as professionals responsible for helping to determine the best possible placement for hundreds of young writers. They did not take this task lightly.

***

References

Campbell, Elizabeth H. (1991). Composition Teachers Talk About Student Essays (Doctoral Dissertation, University of Cincinnati, 1991). Dissertation Abstracts International, 92-154-55.

Campbell, Elizabeth H. Field Notes (unpublished 1990).