

Can Naive Judges Recognize Improvement in ESL Compositions?

by Janet L. Kayfetz and Laurie J. Blass

Can students and teachers expect improvement in ESL composition classes at different levels? Is this improvement significant enough to be recognized by people who are not teachers of ESL? Can these "naive judges" not only perceive the improvement, but also identify specific areas of improvement? Finally, do the areas identified correspond to the classroom focus? These are the questions that guided our study.

Our questions are those that many ESL composition teachers have asked. Most ESL composition teachers are very involved with the day-to-day details of teaching, such as lesson-planning, paper-correcting, writing classroom materials, having individual conferences, etc. In addition, when we evaluate our students' writing, we tend to be very particular about correctness, allowing little room for imperfection. Because of these factors, and because we have so much contact with the students on a regular basis, we often lose sight of the overall progress of their writing ability and even doubt whether or not they improve at all.

We wondered if someone who did not have to pay attention to all the classroom details, did not see students regularly, was not an expert in grammatical correctness, and in fact knew nothing about ESL, would be able to recognize the overall improvement in students' writing that teachers often miss. We felt that such a person, unencumbered by expert "knowledge" would be able to tell us a great deal about our students' improvement.

DESIGN

The hypotheses advanced for this study are: 1) students improve in composition classes, 2) naive judges can recognize this improvement, 3) naive judges can recognize from among three choices the type(s) of improvement and 4) naive judges will identify the types of improvement as being a) clarity

and b) organization and development of ideas.

The study was conducted twice with four groups of students at four different proficiency levels. These levels, which we have designated as A, B, C, and D, were beginning to advanced. The first time, samples of first and last day compositions were collected from twenty-seven foreign students attending UC Berkeley's eight-week ESL Workshop, Summer 1979 (Kayfetz, Blass, and Cato 1979). The sample was divided into two groups: group B, a beginning group of eleven, and group C, an advanced group of sixteen. The first-day composition was a placement exam and the last-day composition was a final exam given in the seventh week of class. The subject for both compositions was "Discuss some of the important consequences of the current oil shortage." In both cases, the students wrote for one hour. The teaching points in the classroom for both groups during the session were clarity and organization and development of ideas, as well as grammar.

The second time the study was conducted, first- and last-day compositions were taken from two different groups, group A and group D. Group A was a beginning group with a lower proficiency in English than group B. It consisted of nine foreign students from the Intensive English Language Center at the University of Nevada-Reno, Fall 1979. Group D was advanced, and at a higher level than group C. It consisted of fifteen students, most of whom were immigrants attending UC Berkeley's advanced ESL composition course, Fall 1979. This is a required course that is offered during the regular school year. As with groups B and C, groups A and D were given a first-day composition and a last-day composition seven weeks later. The topics for these groups differed from the one given to groups B and C. The

students in group A wrote for forty minutes describing an object for both the first- and last-day compositions. The subject for group D's first-day composition was "Should the government have the power to limit the size of families?", and for the last-day composition, the subject was "Should the government reinstate the draft?". As with groups B and C, the students in group D wrote for one hour. Again, the classroom focus for both groups was on clarity and organization and development of ideas, as well as grammar.

The rest of the procedure was the same for both studies. The first- and last-day compositions of each student were stapled together in sets. Sometimes the first-day composition was on top, and sometimes the last-day composition was. For our records, the first day compositions were marked with an π on the back and the last day compositions were marked with a π .

Next, naive judges were selected to read the sets of compositions. By "naive judge" we mean someone who is not an English teacher, ESL or otherwise. Our judges

ranged in age from twenty-two to sixty, and had various occupations, e.g. one was a secretary, one was an accountant, one was a housewife, two were law students, one was a systems analyst, etc.

Each set of compositions was read by three naive judges. The judges were given a questionnaire for each set on which they gave biographical information and indicated their answers. They were asked to determine which of the two compositions in each set was better. They were also asked to indicate which of the following factors influenced their decision: a) clarity, b) organization and development of ideas, and c) grammar. Before the judges began, it was suggested that they read the compositions as they would read a news article or a memo at work. There was a space at the bottom of each questionnaire for additional comments, and the judges were encouraged to comment on each set.

The judges' responses regarding which compositions were better and the factors indicated as influencing their decisions were then tallied.

TABLE 1
Samples and Naive Judge Decisions

Sample	Total no. of comps. in group	Total no. of comps. judged to be better ⁺		No. of votes received	
		No.	%	2/3	3/3
A ₂	9	9	100.0	0	9
C ₂	16	13	81.3	4	9
B ₁	11	7	63.6	5	2
D ₁	15	9	60.0	7	2
D ₂	15	6	40.0	3	3
B ₂	11	4	36.4	2	2
C ₁	16	3	18.7	3	0
A ₁	9	0	00.0	0	0

A = Beginning; Intensive ESL Program, University of Nevada
 B = Intermediate; ESL Workshop – Summer Session, UC Berkeley
 C = Low-advanced; ESL Workshop – Summer Session, UC Berkeley
 D = High-advanced; ESL Subject A (university requirement), UC Berkeley

1 = Composition written first

2 = Composition written seven weeks later.

+ Represents degree of naive judge agreement.

RESULTS

The results were more revealing than we had anticipated. We had expected to find improvement at all four proficiency levels: A, B, C and D. We were interested to note that this did not turn out to be the case. Table 1 summarizes the decisions made by the judges. Specifically, according to the judges, there was significant overall improvement in groups A and C, while the judges could not detect significant improvement in groups B and D. For group A, nine out of nine last-day compositions (100%) and for group C, thirteen out of sixteen last-day compositions (81.3%) were judged to be better. These results were what we expected. For group B, four out of eleven last-day compositions (36.4%) and for group D, six out of fifteen last-day compositions (40%) were judged to be better. These results were unexpected, but they caused us to take a closer look at possible explanations.

Table 2 summarizes the types of improvement indicated by the judges. We expected that the judges would identify the types of improvement as being a) clarity and b) organization and development of ideas since they were the teaching points.

This is exactly what happened. In all cases but two, groups B₁ and B₂, the total of the percentages for a) clarity and b) organization and development was higher than that for grammar, indicating that the students showed greater improvement in those areas. Even in groups B and D, where overall improvement at the end of the course did not seem to occur (according to the judges), the judges chose the better compositions because of their a) clarity and b) organization and development.

DISCUSSION

The present study allowed us to look at the broad continuum of proficiency in ESL composition, from a very beginning level (group A) to a very advanced level (group D). We offer the following as possible explanations for the results obtained for each group.

Group A (Very beginning level)

Because this group was at the very basic proficiency level in second language writing ability, the writing ability had nowhere else to go but up. This explains the unanimous agreement of the judges in selecting nine out of nine last-day compositions as being better. The judges further agreed that all three skill areas, a) clarity, b) organization

TABLE 2
Summary of Factors Influencing Naive Judge Decisions

Sample	Total no. of comps. judged to be better out of group total	Total no. of factor judgments	Clarity		Organization & Development		Grammar	
			No.	%	No.	%	No.	%
A ₂	9/9	33	11	33.3	13	36.4	10	30.3
C ₂	13/16	56	8	14.3	22	39.3	26	46.4
B ₁	7/11	14	1	7.1	5	35.7	8	57.1
D ₁	9/15	30	5	16.7	12	40.0	13	43.3
D ₂	6/15	24	5	20.8	9	37.5	10	41.7
B ₂	4/11	8	1	12.5	2	25.0	5	62.5
C ₁	3/16	11	3	27.3	3	27.3	5	45.5
A ₁	0/9	0	0		0		0	

- A = Beginning; Intensive ESL Program, University of Nevada
- B = Intermediate; ESL Workshop – Summer Session, UC Berkeley
- C = Low-advanced; ESL Workshop – Summer Session, UC Berkeley
- D = High-advanced; ESL Subject A (university requirement), UC Berkeley

1 = Composition written first
2 = Composition written seven weeks later.

and development of ideas and c) grammar showed almost equal improvement and thus made the last-day compositions the better ones. The fact that the judges observed such "uniform" improvement at this level seems to correspond to the observation that many ESL teachers enjoy teaching a very basic level class. The same reasons apply: teachers like to see the obvious improvement in students' abilities, which in this case is very apparent after just seven weeks of instruction.

Group B (Intermediate level)

Group B's level of ability was higher than that of Group A. While the results for group A are quite clear, they are less so for group B: the judges chose only four out of eleven of the last-day compositions (36.4%) as those showing the most improvement. We suggest that a seven week period of instruction at this intermediate level has built into it the obvious limitation of time and the less obvious one of scope. It seems that, for students at this level, exposure to the range of possibilities in English writing may serve to inhibit their performance by confusing them. Furthermore, this confusion may be compounded by an increased awareness of the range of possible

Janet L. Kayfetz is a senior scholar at the Graduate School English Language Center, Beijing, the People's Republic of China, a joint ESP program of the University of California, Los Angeles and the Chinese Academy of Sciences, where she coordinates the speaking and writing curricula. She holds a doctorate in foreign language education with a concentration in adult second language acquisition from Florida State University.

errors they may make. So, whereas many students felt relatively self-confident when they wrote their first-day compositions, these same students felt less confident when asked to write on the same topic for their last-day compositions, especially as this was also part of their final exam.

The observations made here lend support to the feeling of frustration that many ESL

teachers feel when working with students at the intermediate level. While it often appears that students' writing reaches a plateau, the evidence here is that it actually seems to "degenerate" for a period of time.

At this point in the discussion it seems useful to advance an additional explanation for the results observed thus far. The acquisition/learning concepts of Monitor Theory (Krashen 1977) seem to be quite useful in describing what happens as students progress from the beginning level in

Laurie J. Blass is an associate in English as a second language at the University of California, Berkeley, and an Instructor of ESL at UC Extension. She received a masters degree in linguistics with a concentration in teaching English from the University of Pittsburgh. Her interests include the classroom applications of current issues in applied linguistics research.

second language writing ability to the intermediate level. Beginners are dealing with the second language in terms of understanding and communicating a message. Because their knowledge is so limited, they can give little attention to details of grammatical correctness and organizational form. They are *acquiring* the second language itself while at the same time working with the problems of expressing this new language in writing, a task which some say is equivalent to acquiring a completely different system (Keenan 1977, Krashen 1978).

As students acquire enough to gain a foundation in the new language, and understand enough of it to communicate effectively, they can then turn to some of the details of correctness and form. That is to say that they can *learn* rules that can be applied to their writing, such as grammar rules, rules of paragraph development, rules of rhetoric, etc. An over-emphasis on the importance of correctness and rule application, however, can often inhibit students' ability to communicate, since it is difficult to concentrate on both the clarity of the message and its correctness at the same time (Burt and Dulay 1978). In fact, it has been observed

that some students "over-monitor," that is they think so much about correctness that their language is stilted (Krashen and Pon 1975) and can even be less accurate than it is when there is less focus on correctness and more on natural communication of the message (Kayfetz Fuller 1978). The intermediate level, group B, in this study, seems to be at the point where they are beginning to *learn* some of the rules of form, but are not yet able to apply these rules so that their ability level is "boosted." As was mentioned above, they were probably more confused than helped, which explains their apparent regression in ability. Such a group would benefit from more *acquisition*, i.e. communication without an extreme emphasis on correctness.

Group C (Low-Advanced level)

This group's proficiency level was in the lower half of what we would call the advanced level in writing ability. These students had a good command of grammar and a fairly wide range of vocabulary. These compositions were also longer than those of groups A and B.

The judges were impressed with the improvement of these students: they chose thirteen out of sixteen of the last-day compositions (81.3%) as being better. They further indicated that in 53.6% of the cases, these choices were due to either a) clarity of the main idea or b) organization and development.

In terms of the explanation advanced thus far, Group C seems to occupy an exciting place in the continuum of second language writing proficiency. This seems to be the point at which the students are beginning to digest the knowledge that helps to improve writing skill. That is to say that these students have *acquired* enough of the second language to feel a sense of security in their ability to communicate so that some attention can be given to the rules, etc. they have *learned*. This application of some of the rules of correctness, paragraph development, rhetoric, etc. "boosts" the ability level of the writing.

ESL teachers working with students at this level predictably enjoy themselves. They can see the fruits of their labors as

they watch their students' writing improve. This study has shown that even a seven week period of time is long enough for such improvement to occur.

Group D (High-advanced level)

Group D's proficiency level was in the upper half of the advanced level in writing ability. Since most of the students in this group were immigrants attending an American university, their acquaintance with the language and culture could be said to be in many ways more "intimate" than the other groups.

The judges seemed to agree that there was no clear indication of overall improvement after seven weeks of instruction: they chose nine out of fifteen first-day compositions (60%) as being better.

A reasonable explanation for the results for group D is as follows: Students at this advanced level of writing ability have a very good command of the language and have mastered many of the rules of good writing. Changes in ability at such advanced levels are usually subtle and very often occur from actually writing more, without classroom instruction. It seems to us that, in order for significant improvement to occur in a seven week period, students would have to make a conscious commitment to their own progress. That is, they would have to be sufficiently motivated to want to write, rewrite, edit, self-correct, seek help, etc. to boost their already advanced writing ability.

Many of our colleagues who teach this level of ESL composition experience a frustration and fatigue that is almost unavoidable unless their students are highly motivated and self-directed. However, while classroom teaching may be frustrating, one-on-one tutoring often seems quite successful in boosting students' writing ability at this level.

REFERENCES

- Burt, Marina K. and Heidi C. Dulay. 1978. Some guidelines for the assessment of oral language proficiency and dominance. *TESOL Quarterly* 12. 177-92.

Kayfetz Fuller, Janet. 1978. An investigation of natural and monitored morpheme difficulty orders by non-native adult students of English. Doctoral Dissertation, Florida State University.

Kayfetz, Janet L., Laurie J. Blass and Julia Cato. 1979. Naive judges for compositions. *CATESOL Newsletter* 11.

Keenan, Elinor. 1977. Why look at unplanned and planned discourse. In E. Keenan and T. Bennett (eds.), *Discourse Across Time and Space. Southern California Occasional Papers in Linguistics* 5. Los Angeles: University of Southern California, 1-41.

Krashen, Stephen D. 1977. Some issues relating to the Monitor Model. In H.D. Brown, C. Yorio and R. Crymes (eds.), *On TESOL '77*. Washington: TESOL, 144-58.

Krashen, Stephen D. 1978. On the acquisition of planned discourse: Written English as a second dialect. Claremont Reading Conference.

Krashen, Stephen D. and P. Pon. 1975. An error analysis of an advanced ESL learner: The importance of the monitor. *Working Papers on Bilingualism* 7: 125-29.

A paper presented at the Fourteenth Annual TESOL Convention, San Francisco, March 1980.

Author's note: We thank Steve Krashen for his helpful comments and support and June McKay for her comments on an early draft of this paper. Additionally, we would like to thank our colleagues at UC Berkeley who cooperated with us on this project.