Effect of Videotape Replay on the Quality and Accuracy of Student Self-Evaluation

PHILIP B. PALMER,
JILL NEWMAN HENRY,
and DAVID A. ROHE

This study was designed to determine the effect of videotape replay on the quality and accuracy of student self-evaluation. Thirty-two junior physical therapy students were asked to perform a clinical skill and to assess their performances using the Performance Evaluation Form (PEF) designed for this study. The students in the Experimental Group assessed their performances after viewing a videotape replay. Students in the Control Group assessed their performances without viewing videotape replays. Faculty judges used the PEF to rate each student's performance. The instructor and student ratings were compared to determine the accuracy of student ratings. In addition, two faculty judges used a scale designed for this study to rate the degree of quality of each student's self-evaluation. No significant differences were found between the Experimental Group and the Control Group on either of the variables tested. The data gathered in this study indicated that videotape replay lacks value in enhancing self-assessment when evaluation immediately follows the performance of a skill. Further study is indicated.

Key Words: Educational measurement, Feedback, Physical therapy, Students.

The ability to evaluate one's own performance has been described as a "hallmark of professionals."1 Although professionals are expected to assess their own performances, instruction in self-evaluation is rarely a part of their training. Instead, the assumption is made that the professional will develop the necessary skills for effective self-evaluation after graduation. Muse warns against this practice: "Competence in learner self-evaluation is not to be acquired by reiteration of its necessity by teachers and postponement of initial practice until after graduation."2 This warning and the belief that the ability to assess one's own performance can be improved with practice have led several authors to suggest that training in self-evaluation be incorporated into school curricula.3,4

The faculty of the Department of Physical Therapy at the Medical College of Georgia is seeking the most effective methods to instruct students in self-evaluation and to facilitate the development of self-evaluative skills. Videotape replay (VTR) of the students' performances on clinical practical examinations is used to facilitate more realistic and accurate self-evaluations. Little empirical evidence exists to validate the belief that this time-consuming process of recording and reviewing the students' performances does indeed result in more accurate or realistic self-evaluations of those performances. The purpose of this study, therefore, was to determine the effects of VTR on the quality and accuracy of student self-evaluations. If VTR increases the quality and accuracy of self-evaluation, its use should be encouraged. If, however, VTR has no effect on these variables, the use of a less costly and less time-consuming process should be encouraged.

REVIEW OF LITERATURE

A review of literature did not reveal any articles that specifically addressed the issue of the effects of VTR on self-evaluative skills. Three articles were found, however, that supported the belief that VTR is a potentially effective aspect of self-evaluation.5-7

Saarinen and associates described the use of video feedback in the evaluation of the clinical performance of physical therapy students.6 In a program started in the spring of 1975, students were provided with VTR of their physical therapy sessions with simulated patients to identify strengths and weaknesses. One of the primary purposes of this program was to assist in the development of self-evaluation skills. The authors found that the accuracy and honesty of student self-evaluation improved with subsequent sessions of video feedback. This study did not use a control group and, therefore, it is impossible to determine if the improvement was a result of the use of VTR or of practice in self-evaluation.

Although not related to the use of VTR in facilitating self-evaluation, we found two articles that supported using VTR to provide feedback to students. In a study on the effects of using VTR on student-teacher interaction, Davis and Dans concluded that the use of video feedback resulted in greater specificity and depth of feedback given the students concerning their performances.8 Schmidt and Messner lent support to this conclusion by stating that the use of videotape playback...
The judges' scores were correlated using the Pearson product-moment correlation to determine the coefficient of interrater reliability. Therefore, we concluded that the potential for internal invalidity based on instrumentation had been reduced.

### Procedure

Thirty-two junior physical therapy students at the Medical College of Georgia were randomly assigned to either the Experimental Group or the Control Group. Each student was presented with a problem requiring use of manual muscle testing and goniometry skills and asked to demonstrate those skills on a simulated patient. Each student's performance was videotaped. Immediately after the session, the student was asked to complete the PEF designed for this study. The Experimental Group completed the form after viewing the videotapes. The Control Group completed the form before viewing the videotapes. Students in the Control Group were not allowed to make any changes on the PEF after viewing their videotape recordings.

The videotapes of the students' performances were reviewed by two faculty members who used the PEF to rate the performance of each student. The instructors' ratings of each student's performance were compared with the student's self-rating to determine the relationship between the two sets of ratings. We assumed the instructors' ratings were "true." In other words, the instructors knew what the perfect performance would be and could observe and measure every part of the performance by the student. The instructors' ratings were used as the standard to measure the students' ratings for accuracy.

Two faculty judges used the QRS to rate the quality of the students' self-assessments. The scores for each student were averaged and compared to determine differences between the Control and Experimental Groups.

### RESULTS

Table 1 presents the results of the *t*-test we used to determine the degree of difference in the quality of student self-evaluation between the Control and Experimental Groups. No significant difference was found between the two groups and, therefore, the hypothesis concerned with the quality of student self-evaluation could not be accepted.
Table 2 presents the results of the correlations between instructor and student ratings for each group and all students combined. The value calculated using Fisher's $z_t$ transformation was 0.19. This value was less than the value required for significance at the .05 level. Thus, the hypothesis related to the accuracy of student self-ratings could not be accepted. All of the correlations, however, were significantly greater than zero.

**DISCUSSION**

This study was designed to provide empirical evidence to justify the use of videotape playback of students' clinical performances as a way to increase the quality and accuracy of student self-evaluation. Analysis of the data gathered during this study failed to demonstrate significant differences between the Experimental and Control Groups on either of the variables considered. The findings, rather, indicated that the use of procedures that are less time-consuming and less costly than VTR may be more beneficial in providing feedback to students and in facilitating immediate self-evaluation. This study needs to be replicated to verify whether VTR lacks value for self-assessment immediately after performing a skill.

Research cited earlier claimed that video feedback resulted in greater specificity and depth of feedback given students concerning their performances. We assumed that video feedback would have a similar effect on the quality of feedback the students gave themselves during the self-assessment process. Thus, the students who had the opportunity to review videotape recordings of their performances would generate self-evaluations of a higher quality than students who did not have that opportunity. In this study, however, we found no significant difference between the Experimental and Control Groups related to the quality of student self-evaluation. We also assumed that greater specificity and depth of feedback would result in more accurate self-ratings. In other words, the correlation between student and instructor ratings would be higher for students receiving video feedback because the quality of self-evaluation would be greater. Again, the results showed no significant difference between the two groups on the correlation between the student and instructor ratings. Two reasons for the lack of significant differences between the two groups on both variables are plausible. Each has implications for the use of VTR to facilitate self-evaluation.

First, the students participating in this study were asked to assess their performances immediately after the completion of the practical examination. Thus, the ability to recall specific incidences in which the students had performed well or poorly was enhanced. Review of the videotape only served to confirm the students' assessments of their performances. When students do not have the opportunity to assess their performances immediately after the completion of the performances because of time constraints or other factors, VTR might be more effective in facilitating greater accuracy and quality in self-evaluation. Video feedback would serve to refresh the students' memories and to help them recall specific behaviors during the performance of clinical skills. In such a situation, VTR should have a positive effect on the quality and accuracy of student self-evaluations. Further research should be done to determine the effects of VTR on the accuracy and quality of self-evaluation when assessment is delayed.

Second, the failure to demonstrate significant differences between the Experimental and Control Groups in this study may be related to the lack of experience of the Experimental Group in using video feedback. This study was the first opportunity for the students to complete a formal self-assessment on their performances of a clinical skill. In addition, it was the first time that most of the students had observed themselves on videotape. Research indicates that the ability to perform self-evaluation is increased when students are repeatedly asked to evaluate their own performances and that honesty and accuracy in self-assessment are improved with repeated opportunities to review subsequent performances on videotape. In essence, students must learn to use videotape replay effectively as a source of feedback and to become more effective self-evaluators. Research to determine the effects of instruction on the use of VTR as a source of feedback in self-evaluation needs to be carried out.

Some educators are skeptical of the use of self-evaluation because of the low correlations between self-ratings and ratings by others. The correlations between student and instructor ratings in this study were significantly greater than zero (Tab. 2) and indicated a high degree of accuracy on the part of the students in rating their performances. This finding contradicts the data found in the literature. Several studies indicated that correlations resulting from comparison of self-ratings and ratings by others were low and ranged from .19 to .32.12,13

For other educators, self-evaluation may be suspect because of the belief that students have a tendency to inflate self-ratings. Research indicates, however, that self-ratings are generally lower than ratings by others.14,15 This study supports that research. Except for a few cases, the students rated themselves lower than either of the instructors.

A general belief is that students may inflate ratings if self-evaluation is used for making decisions regarding their performances or for the purpose of grading.5,16 Woodward claims that if self-evaluation is done for grading, "we may teach self-deception, rather than self-evaluation."16 The students in this study were self-evaluating in a nongraded situation; therefore, the self-ratings are more likely to be accurate assessments of their performances.

Despite these problems, self-evaluation is important, and its use is to be encouraged. Self-evaluation facilitates greater self-awareness,17 helps to develop the skills necessary for self-appraisal,15 and provides motivation for change to occur.2 In fact, all evaluation must become self-evaluation if it is to be assimilated.

**CONCLUSIONS**

Because professionals are expected to assess their own performances, instruction in self-evaluation must be incorporated into professional curricula. Claims have been made concerning the effectiveness of VTR in facilitating self-evaluation. The findings in this study, however, failed to support these claims and indicated a need to investigate further the effectiveness of videotape playback in developing self-evaluative skills. In particular, the need exists to determine when the use of video feedback is most effective for facilitating self-appraisal and when a less time-consuming and less costly procedure may be used to accomplish this goal. The data gathered in this study indicate that VTR lacks value in en-
hancing self-assessment when evaluation immediately follows the performance of a skill. This study needs to be replicated to verify this finding. Further study should also be carried out to determine the effect of delayed and immediate VTR, as well as the effect of instruction in the use of VTR as a source of feedback, on self-evaluative skills.

Acknowledgments. We thank Patricia Hanner for her assistance in videotaping the students' performances, Harry C. Davis for his assistance in data analysis, and Kenneth Wesling, PhD, for his critical thoughts on this manuscript.

REFERENCES

5. Willoughby L: Self-evaluation in Medical Education at the University of Missouri-Kansas City. In: Research in Medical Education: Proceedings of the Twentieth Annual Conference of the Association of American Medical Colleges, Division of Educational Measurement and Research. Washington, DC, November 1981, p 236

APPENDIX 1
Examples of Items from the Performance Evaluation Form

II. Rated elements: For items in section A, Yes = 1 and No = 0; for items in sections C and D, Yes = 5 and No = 0. For quality dimensions, 5 = excellent, 1 = poor, and 0 = did not perform. If an item is not applicable, write NA next to the item and do not score it. The final score will be adjusted to compensate for this. Cite examples that justify your ratings in the space provided to the right of each item.

A. Communications
1. Introduces self (name and title)
2. Demonstrates desired motions
3. Quality dimensions
   a. Explanation of procedures
   b. Assessment of patient's understanding (eg, asking questions)
   c. Rapport with patient
D. Goniometry
1. Tasks
   a. Requires patient to perform, or passively performs, movement several times before measuring
   b. Places arms of goniometer correctly, as described in class
   c. Reads goniometer correctly
   d. Records or reports measurements correctly
   e. Checks same movement on opposite side
   f. Able to visualize or palpate bony landmarks

Quality dimension
2. Smoothness of application (manner in applying individual components of test)

E. Overall performance
1. Ability to modify based on unanticipated responses
2. Efficiency of performance (movement from one component of the process to another, eg, intro. to MMT to ROM)

III. Additional questions to be used in self-evaluation.
A. What did you do well on the practical?
B. What would you do differently if you had the chance?
C. Identify areas in which you need further practice or information.
D. How did you feel during the practical? What factors contributed to your feelings? Identify both positive and negative factors if they apply.
### APPENDIX 2

Examples of Items from the Quality Rating Scale

(Note: 0 = no examples, behaviors, or factors cited)

1. **Degree of specificity of examples cited to support ratings**
   - 0 = general
   - 1, 2, 3, 4 = specific

2. **Degree of relationship of examples to items for which they are cited**
   - 0 = no relationship
   - 1 = examples relate well to items
   - 3, 4, 5 = all extreme ratings verified by examples

3. **Degree of completeness of cited examples**
   - 0 = extreme ratings not verified
   - 1, 2, 3, 4 = all extreme ratings verified by examples

4. **Degree of specificity of behaviors identified in**
   - a. **Question 1**
     - 0 = general (e.g., good communication)
     - 1, 2, 3, 4, 5 = specific (e.g., clear, concise explanation)
   - b. **Question 2**
     - 0 = general (e.g., MMT)
     - 1, 2, 3, 4, 5 = specific (e.g., stabilize, check opposite side)
   - c. **Question 3**
     - 0 = general (e.g., more practice)
     - 1, 2, 3, 4, 5 = specific (e.g., practice applying resistance)

5. **Degree of relationship between factors and feelings identified in Question 4**
   - 0 = no relationship
   - 1, 2, 3, 4, 5 = factors relate well to feelings

\[
\begin{align*}
\text{# examples cited} &= \frac{28}{\text{(rating)}} \\
\text{# examples cited} &= \frac{28}{\text{(rating)}}
\end{align*}
\]