Research Report

Critical Appraisal of Research Literature by Expert and Inexperienced Physical Therapy Researchers

Elizabeth Domholdt
Julianne Lonz Flaherty
James M Phillips

Background and Purpose. *In this study, critical appraisal skills of physical therapists who were inexperienced and expert researchers were compared and the effect of the following independent variables on the appraisal skills of novices was examined: years of physical therapy practice, postgraduate experience with research, level of comfort with research, journal reading habits, and content-specific expertise.* ***Subjects and Methods.*** *Four expert and 20 inexperienced researchers critiqued the same research article. A content analysis of the expert critiques was used to develop a set of 12 items of concerns about the study.* ***Results.*** *Experts identified a mean of 7.75 items of concern; novices identified a mean of 3.95 items. For 10 of the 12 items, a greater proportion of experts identified each concern compared with the novices. Novices had comparatively less difficulty identifying concerns with internal validity than they did identifying issues relating to construct validity. Only one of the variables—content-specific expertise—was associated with differences in critical appraisal skills among the novices.* ***Conclusion and Discussion.*** *These findings suggest that analysis of construct issues in research articles is a more advanced skill than analysis of research design and that critical appraisal of research literature is enhanced by clinical experiences related to the subject of the research report. [Domholdt E, Flaherty JL, Phillips JM. Critical appraisal of research literature by expert and inexperienced physical therapy researchers. Phys Ther. 1994;74:853–860.]*

Key Words: *Physical therapy profession, professional issues; Publications and audiovisual materials; Research.*

Critical appraisal has been described as the "ability to assess the validity and applicability of clinical, paraclinical, and published evidence and to incorporate the results of this assessment into patient management."[1] Many physical therapists today believe that the ability to critically appraise the profession's research literature is a key to the advancement of the profession and science of physical therapy. The accreditation criteria established by the Commission on Accreditation in Physical Therapy Education (CAPTE) provide evidence of this belief in the statement that "program graduates are able to apply basic principles of the scientific method to read and interpret professional literature, to participate in clinical research activities, and to critically analyze new concepts and findings."[2(p14)] Rothstein, as editor of *Physical Therapy*, has stated his belief in the importance of research evaluation by expressing concern that "some therapists often uncritically

E Domholdt, EdD, PT, is Associate Professor and Dean, Krannert School of Physical Therapy, University of Indianapolis, 1400 E Hanna Ave, Indianapolis, IN 46227-3697 (USA). Address all correspondence to Dr Domholdt.

JL Flaherty, PT, is Physical Therapist, Riverview Hospital, Noblesville, IN 46060.

JM Phillips, PT, is Physical Therapist, Beverly Enterprises, Indianapolis, IN 46032.

This research was completed in partial fulfillment of the requirements for Ms Flaherty's and Mr Phillips' Master of Science degree, Krannert School of Physical Therapy, University of Indianapolis.

This study was approved by the Committee on Research Involving Human Participants at the University of Indianapolis.

This article was submitted November 16, 1992, and was accepted April 19, 1994.

accept all that they read in this journal."[3(p278)]

Despite these beliefs in the importance of critical appraisal of the literature, we could locate only one study that assessed the research critiquing abilities of physical therapists. This study, a 1990 survey of the American Physical Therapy Association (APTA), demonstrated that only 50% of surveyed members felt that they had "sufficient working knowledge of research methodology and design enabling them to read journal articles."[4(p69)]

Although the critical appraisal skills of physical therapists are not well documented, several researchers have studied this ability in medical students or residents. Most of the studies are experimental in nature, making either pretest-posttest comparisons with a single group who received a certain instructional program designed to improve critical appraisal skills or comparisons between groups receiving different kinds of instruction in critical appraisal.

The variables of interest within these studies have included reading habits,[5,6] knowledge of epidemiology and biostatistics,[5,6] and ability to critically evaluate a particular research article. The tests of ability range from examination questions developed for specific articles,[7] directed critiques that ask students to assess particular aspects of a given article,[6,8,9] and an all-or-none test of whether students noticed that the methods section of a different article had been inserted into the article of interest.[10]

The results of the studies of critical appraisal skills are ambiguous. One study demonstrated no differences in critical appraisal skills between members of a journal club and a control group[6]; one showed a trend toward improvement of critical appraisal skills after completion of a computer-aided instruction program[8]; one indicated that second- and third-year residents who had completed an epidemiology seminar series had better critical appraisal skills than first-year residents who had not completed the series[7]; one showed that critical appraisal skills improved more when medical students were taught by faculty tutors with training in clinical epidemiology compared with untrained faculty tutors[9]; and one revealed that only a disappointing one of eight students detected a substitute methods section within a research article.[10] Additional reports of journal club activities for other professionals such as social workers[11] and nurses[12] have focused on broadly defined effects on professional development (eg, building networks with colleagues and becoming more aware of professional issues) rather than on specific improvements in critical appraisal skills.

Apart from the inconsistent results about changes in critical appraisal skills after various instructional methods, we find two major limitations within the literature on critical appraisal skills of medical students and residents. First, all the measures of critical appraisal skills have provided subjects with a framework for the analysis of articles. Thus, these authors have not studied critical appraisal skills in the manner in which practitioners will eventually assess literature—that is, without the cues provided by a framework furnished by researchers. Second, although several studies used expert researchers to provide a standard against which the responses of novices could be assessed, none examined the content of the experts' and novices' critiques to determine the nature of the differences in their critical appraisal skills.

Given the lack of information about critical appraisal skills of physical therapists and the limitations of the literature on critical appraisal skills of medical students and residents, we designed a study with the following purposes:

1. To compare the content of research critiques written by expert and inexperienced researchers. We assumed that experts would be more skilled than novices and were interested in documenting the ways in which the novices' skills differed from those of the experts.
2. To determine whether any of the following factors explain differences in novices' ability to critique a research article: years of physical therapy practice, postgraduate experience with research, level of comfort with research, journal reading habits, and content-specific expertise. We hypothesized that novices with more years of clinical practice, more postgraduate experience with research, a higher perceived comfort level with research, and more frequent journal reading habits would be more effective appraisers of the research article selected for critique within this study. The final variable—content-specific expertise—was not one we had initially planned to study. We added the variable during the course of data analysis when we observed that the quality of the critiques seemed related to the area of practice of the respondents.

Method

Subjects

Expert physical therapy researchers were defined as those who regularly serve as reviewers or editors for rehabilitation journals. Five professionals were asked to serve as expert researchers. These professionals were chosen because of their extensive experience reviewing and evaluating research in physical therapy. Four of five experts (80%) who were contacted agreed to participate in this study. These experts had an average of 15 years of experience as physical therapists and an average of 6 years of experience as journal article reviewers or editorial board members. They all possessed either academic or professional doctoral degrees.

Inexperienced physical therapy researchers were defined as clinicians who have had academic exposure to research but who have not made research activities a major focus of their professional lives. We selected

novices by using a sampling frame of current students or graduates of the 1982 through 1992 professional program classes at our institution. All had either completed or were in the process of completing a research course and project while students in the professional physical therapy program. Seventy-five subjects were randomly selected from eight classes of graduates and three classes of enrolled students. Thirty-one of 75 inexperienced subjects (41.3%) agreed to participate. Forty-four inexperienced subjects (58.7%) either did not respond or chose not to participate. Twenty of the 31 novices (64.5%) who agreed to participate actually did so, for a total participation rate of 26.7% (20 of 75) for the novices. The most frequently cited reason for participation was a commitment to the university.

The characteristics of the novices formed the subgroups that we used to determine whether there were differences in novices' ability to critique the literature. The subgroups are based on years of practice, postgraduate experience with research, level of comfort with research, journal reading habits, and content-specific expertise. Eleven novices had 3 or fewer years of clinical experience; 9 novices had more than 3 years of experience. Fourteen novices had had no research experiences after graduation; 6 novices had limited experiences, primarily as data collectors. Fourteen novices read fewer than five journal articles monthly; 6 novices indicated that they read five or more articles per month. Novices were evenly divided on the comfort-with-research variable; 10 novices had neutral to high levels of self-reported comfort ("neutral comfort" to "complete comfort," coded 3 through 5 on a five-point Likert-type item), and 10 novices had low levels of self-reported comfort ("not at all comfortable" to "somewhat uncomfortable," coded 1 and 2 on the Likert-type scale). Eleven novices worked in settings in which it was unlikely that patients would be rehabilitated after anterior cruciate ligament (ACL) reconstruction; 9 novices worked in settings in which such content-specific experience was likely.

Procedure

Draper's[13] study of electromyographic biofeedback and recovery of quadriceps femoris muscle function following ACL reconstruction was used to determine respondents' ability to critique clinical literature. This article was selected for two reasons. First, it dealt with a relatively common clinical presentation with which most physical therapists would have had experience either as a student or a professional. Second, the very nature of the research design—a clinical study in which each subject was assessed over a 3-month period—exposed it to the types of threats to internal validity (such as difficulty controlling outside activities and attendance at therapy sessions) we hoped subjects would identify.

The 5 experts and 75 novices were mailed or given letters of information about the research with a request for participation. Those who agreed to participate were sent a photocopy of the Draper article[13] and instructions to complete a free-form critique, that is, to generate their own comments, concerns, and conclusions about the article without recommendations on structure or format from the researchers. The experts were requested to return their curriculum vitae with the completed critique. The novices were asked to complete a brief questionnaire about employment history in physical therapy, research experience since graduation, reason for participation in the study, journal reading habits, and level of comfort with research. A follow-up letter and an additional copy of the article were sent to novices who had not returned a critique 1 month after the first mailing. A follow-up questionnaire was also sent to each novice who did not complete a critique. This questionnaire addressed issues of comfort with research, journal reading habits, and reasons for nonparticipation. Twenty-nine of the 44 original nonparticipants (66%) returned the follow-up questionnaire. The most frequent reason for nonparticipation was lack of time due to personal or work commitments.

Data Analysis

Participant/nonparticipant response bias. We performed a *t* test and a chi-square test to determine whether there were differences between respondents and nonrespondents on the variables of comfort with research and journal reading habits, respectively. The alpha level of significance was set at .05 for each analysis.

Content analysis. The content of the free-form critiques was analyzed by the three investigators. Each of us read the expert critiques and developed a system of categorizing the content. Then we met to merge our independent categorization systems into a single mutually acceptable system. This categorization consisted of a total of 12 items of concern about the research article. This system was then applied individually by each of us to categorize the responses from the novices' critiques. The Table shows the final 12 items that were used to categorize all of the experts' and the novices' critiques, along with samples of the types of statements considered representative of each category. After developing the categorization system, we realized that 10 of the 12 concerns fell within more general categories of internal, external, and construct validity, as described by Cook and Campbell.[14] A fourth general category was related to the manuscript itself, and not to the research. The list of items in the Table is subdivided into these four general areas of concern. We did not attempt to categorize the seriousness of the concerns or to examine whether either group was better at identifying the most critical threat to validity.

The reliability of our categorization of the novices' critiques was fair on first comparison of our individual results; we were in complete agreement on 67% (60 of 91) of the concerns identified. With discussion, we came to consensus on the other 33% (31 of

Table. *Respondent Statements Representing Each Item of Concern*

Item	Statement
Internal validity	
Poor control of postoperative treatment or outside activity level	"Did you collect any data on daily routine of patients for time spent away from the clinic?"
Poor control of preoperative or surgical differences	"There is no information about the premorbid condition, habits, and activity level of the subjects."
Concerns about lack of accounting for postoperative complications	"No discussion of or allowance for the variables of edema and pain."
External validity	
Comment on size and characteristics of sample	"Generalized too quickly from sample of 22 to the population."
Concern over possible investigator bias	"Possible limitation—bias of data collector—not blind [to group membership]."
Construct validity	
Concerns about presumed relationship between strength and range of motion	"Whole concept of biofeedback that facilitates muscle contraction having anything to do with return of range of motion is lost to me."
Concerns about technique of strength measurement	"A number of published studies cast doubt on the ability to reach peak torque in 3 seconds."
Concerns about amount and type of feedback	"I suggest that the comparison in this study is on adequate versus inadequate feedback."
Concern about poorly developed theoretical framework	"Neither review of literature nor the discussion raise an issue that is well-documented (ie, the selective atrophy of Type I muscle, which may also relate to the training intervention of isometric exercise)."
Concern about lack of functional measures	"Some type of perceived disability scale could also have been useful in determining whether the patients' function was progressing as nicely as their strength and range of motion."
Manuscript concerns	
Missing elements	"The question of why torque values were not gravity corrected should be addressed if it is important enough to mention."
Mislabeled axis in the figure	"Vertical axis is labeled as peak torque, but I think it really shows the mean of peak torque ratios (percentage) at each angle."

91) of the items. To provide a concise summary variable, we then totaled the number of items of concern identified in each critique.

Assessment of differences in critiquing ability. Because of the small sample size, and particularly the small size of the subgroups, we chose not to use statistics to analyze differences in critiquing ability among the subgroups. Instead, tables and graphs of data were generated and analyzed visually. We arbitrarily determined that a 1.0-item difference among subgroup means for the number of items identified or a greater than 10% difference in subgroup percentages for the individual items was needed before we would identify a trend. For example, if the mean numbers of items identified by two subgroups were 3.0 and 4.2, we identified a trend in favor of the second subgroup because the difference exceeded our 1.0-item criterion. If 40% of the respondents in one subgroup identified a particular item of concern and 49% of the respondents in another subgroup identified the same item of concern, we did not consider this modest difference to be important because it failed to exceed our 10% criterion.

Results

Participant/Nonparticipant Response Bias

Nonparticipants actually listed a higher level of comfort for research than did participants (nonparticipant mean comfort score=3.2, participant mean comfort score=2.6, $t=1.8299$, $P=.0352$). No difference was demonstrated in the proportion of participants versus nonparticipants who read either fewer than five articles per month or five or more articles per month (participants=70% and 30%, respectively; nonparticipants=61% and 39%, respectively; $\chi^2=1.9098$, $P=.167$).

Subgroup Comparisons

Novice/expert comparison. Experts identified a mean of 7.8 items of concern; novices identified a mean of 3.9 items. Figure 1 shows the individual item results for the novice/expert comparison. For 10 of the 12 individ-

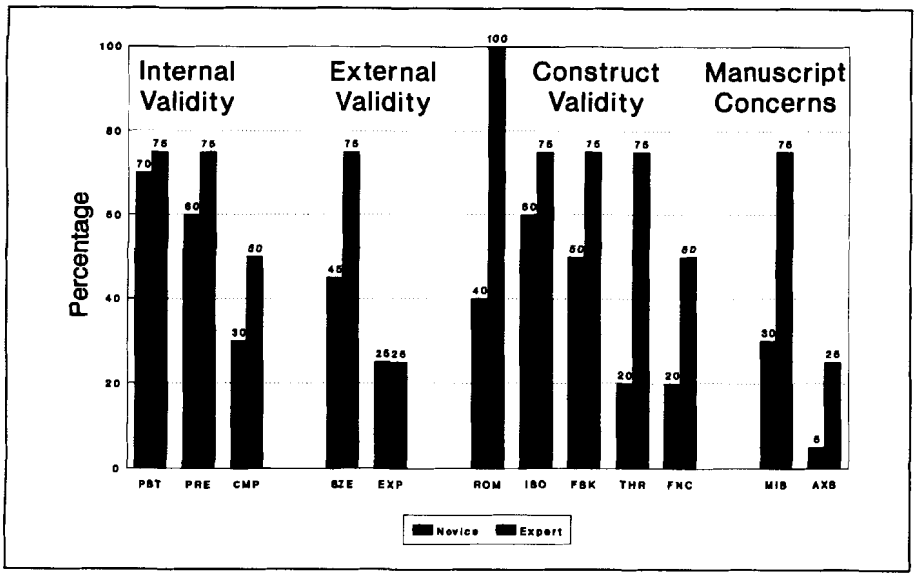

Figure 1. *Percentages of novices and experts who identified each individual item of concern, grouped by types of validity. PST=poor control of postoperative treatment, PRE=poor control of preoperative differences, CMP=concern about not accounting for postoperative complications, SZE=comment on size and characteristics of sample, EXP=concern about experimenter expectancies leading to bias, ROM=concern about assumed relationship between strength and range of motion, ISO=concerns about technique of isokinetic measurement, FBK=concern about amount and type of feedback, THR=concern about poorly developed theoretical framework, FNC=concerns about lack of functional measures, MIS=missing manuscript items, AXS=mislabeled figure axis.*

ual items, a higher proportion of the experts than the novices identified the concern. For the other 2 items, neither group was favored. The differences in the proportion of experts and novices who identified certain items of concern was least marked for the internal validity items (differences in novice and experts who identified each item ranged from 5% to 20%) and most marked for the construct validity items (differences ranged from 15% to 60%). Both the mean number of items identified and the individual item results support our assumption that experts would provide more complete critiques of the journal article.

Experience in physical therapy. Novices with more than 3 years of experience as physical therapists identified a mean of 4.1 items of concern; novices with 3 or fewer years of experience identified a mean of 3.7 items. For 6 of the 12 individual items, a higher proportion of the more experienced therapists than the less experienced therapists identified the concern. For 3 of the 12 items, neither group was favored, and for the other 3 items, a higher proportion of the less experienced therapists identified the concern. Neither the number of items identified nor the individual item results support our hypothesis that those therapists with more clinical experience would have better critical appraisal skills that those with less experience.

Postgraduate experience with research. Novices with postgraduate research experience identified a mean of 3.6 items of concern; novices without such experiences identified a mean of 4.0 items. For 1 of the 12 individual items, a higher proportion of the therapists with postgraduate research experiences than those without such experience identified the concern. For 7 of the 12 items, neither group was favored, and for the other 4 items, a higher proportion of those without postgraduate research experience identified the concern. Neither the number of items identified nor the individual item results support our hypothesis that those therapists with more postgraduate research experience would have better critical appraisal skills than those with less research experience.

Journal article reading habits. Novices with more regular reading habits identified a mean of 3.6 items of concern; novices with less regular reading habits identified a mean of 4.0 items. For 3 of the 12 individual items, a higher proportion of the more regular readers than the less regular readers identified the concern. For 4 of the items, neither group was favored, and for the other 5 items, a higher proportion of the less regular readers identified the concern. Neither the number of items identified nor the individual item results support our hypothesis that those therapists with more regular reading habits would have better critical appraisal skills than those with less regular habits.

Comfort with research. Novices with a higher level of comfort with research identified a mean of 4.4 items of concern; novices with lower levels of comfort identified a mean of 3.4 items. For 4 of the 12 individual items, a higher proportion of the therapists with higher comfort levels than those with lower levels identified the concern. For 7 of the 12 items, neither group was favored, and for the other item, a higher proportion of those with lower comfort levels identified the concern. Neither the number of items identified nor the individual item results provide strong support for our hypothesis that those therapists with a higher level of comfort with research would have better critical appraisal skills than those with a lower level of comfort.

Content-specific expertise. Novices with positions in which they were likely to treat patients after ACL reconstruction identified a mean of 4.8 items of concern; novices with positions in which treatment of such patients was unlikely identified a mean of 3.3 items. Figure 2 shows that for 8

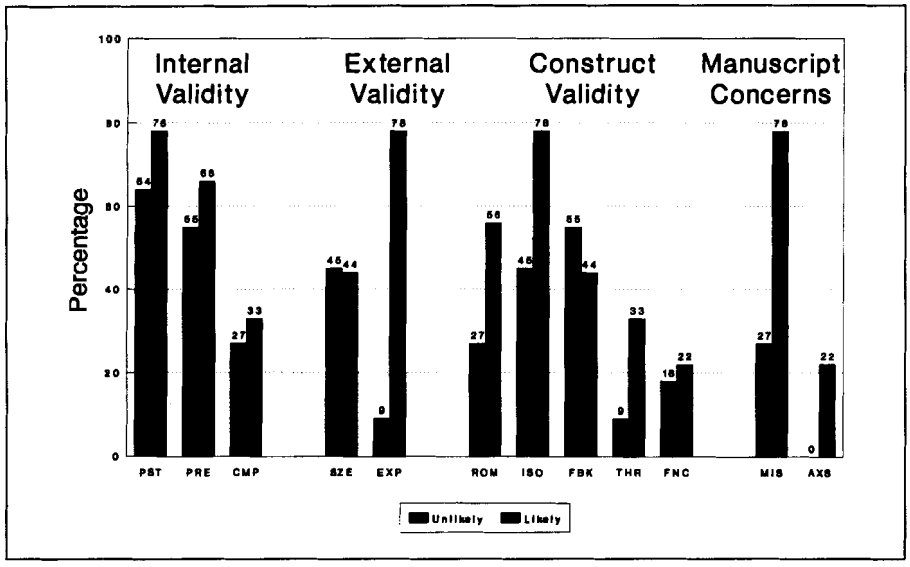

Figure 2. *Percentages of respondents who were likely or unlikely to see patients after anterior cruciate ligament reconstruction who identified each individual item of concern, grouped by types of validity. See Figure 1 legend for descriptions of abbreviations.*

of the 12 individual items, a higher proportion of the therapists who were likely to have content-specific expertise than those who were unlikely to have such expertise identified the concern. For 3 of the 12 items, neither group was favored, and for the other item, a higher proportion of those unlikely to have such expertise identified the concern. Both the mean number of items identified and the individual item results tend to support our subjective impression that those therapists with experiences in settings where they would likely treat patients after ACL reconstruction provided more complete critiques of the journal article than those who were unlikely to treat such patients.

Discussion

As anticipated, we found differences between the critiques of the expert and inexperienced researchers. Specifically, we found that the difference between the percentages of experts and novices identifying construct validity concerns was large, with a much higher proportion of the experts identifying each concern. In contrast, the difference between the percentages of experts and novices identifying the various internal validity concerns was fairly small. Intuitively, it seems that identifying concerns about the structure of a study (a component of internal validity) is a more concrete process than determining concerns about the ideas that underlie a study (construct validity). Our research indicates that the respondents—physical therapists who graduated from one academic institution and were inexperienced researchers—were able to master the former process fairly easily, but had difficulty with the latter process. We did not, however, examine the ability to discern the most critical issues relating to validity for either group. Faculty who teach research, or who integrate critiquing of research reports into their courses, may wish to consider ways in which they might develop the more abstract and theoretical aspects of critical appraisal of research literature in their students.

Only one factor, content-specific expertise, was associated with differences in the critical appraisal skills of novices. Novices who worked in positions in which rehabilitation of patients after ACL reconstruction was likely identified more concerns than did those who were unlikely to treat such patients. In retrospect, it seems clear that the critique process requires that reviewers draw on two domains of knowledge—knowledge related to the research process and knowledge related to the content of the research. We partially acknowledged these two domains in our research design by selecting an article about a common clinical presentation that we assumed most physical therapists would have had experience with either as a student or a professional. The content-specific expertise results, however, indicate that it was unreasonable to expect that therapists working in different settings would identify the same sorts of concerns about the article. For example, the construct validity concerns that dealt with the specific strength and range of motion measures used in the study were more frequently cited by those who were likely to work with patients after ACL reconstruction. Concerns about the sample size also were far more likely to be raised by individuals who are likely to work with patients after ACL reconstruction. Perhaps these therapists were less accepting of the small sample size because they are in positions that make them aware of the large numbers of potential subjects who have undergone ACL reconstruction.

Our conclusions about content-specific expertise must be tempered by the fact that this variable was developed during the process of data analysis. Because we had not anticipated the inclusion of this variable at the start of the study, the experts were selected on the basis of research expertise and not content expertise. Researchers with combined research and content expertise may have provided a different standard of comparison. In addition, the employment data for the novices were not designed to determine whether they treated patients after ACL reconstruction. Therefore, we made assumptions about the likelihood of such treatment on the basis of our available information (for example, novices employed in pediatric settings were placed in the "not likely" group and novices employed

in outpatient orthopedic settings were placed in the "likely" group).

Despite these limitations, educators who teach critical analysis of research—particularly those who teach practicing clinicians—may wish to consider allowing students a selection of article topics to maximize their ability to detect validity problems within the studies. Educators who teach professional program students (those who are receiving their initial preparation as physical therapists) may be able to assume a more uniform level of preparation across the content areas of the discipline because these students have not yet had the opportunity to specialize their practice. Indeed, the distinction between broadly educated students and more narrowly focused practitioners may explain why previous studies of the critical appraisal process have not identified content-specific expertise as a factor of importance. These studies either compared medical students with one another[8–10] or compared residents in the same specialty areas.[5–7] These comparisons do not lead to the differences in practice that were evident in our sample of clinicians.

There were no consistent differences in critical appraisal skills between those therapists with 3 or fewer years of clinical practice and those with more than 3 years of practice. This finding seems to indicate that clinical experience, in and of itself, does not lead to improvements in critical appraisal skills. Therefore, therapists who wish to refine their critical appraisal skills after graduation may wish to seek out a structured program—such as continuing education focused on research or enrollment in graduate studies—to assist them in doing so.

There were also no consistent differences in critical appraisal skills for those therapists with and without postgraduate experience in research. This finding may be related to the low level of postgraduate research experiences of graduates. Most of the six novices with postgraduate experiences had only participated as data collectors, not as primary researchers or authors. Because their limited experiences, for the most part, did not require them to be the primary intellectual contributors to the projects in which they participated, it is not surprising that they did not demonstrate better critical appraisal skills than did their counterparts without postgraduate research experiences.

There were also no consistent differences in critical appraisal skills for those therapists with high and low self-reports of comfort with research. We believe that this finding indicates that self-reported comfort with research is not a particularly useful indicator of critical appraisal skills.

There were also no consistent differences in critical appraisal skills between those therapists who read fewer than five journal articles per month and those who read five or more journal articles per month. This finding is consistent with the findings of Linzer and colleagues.[6] They found that although reading habits of residents improved after participation in a journal club, critical appraisal skills did not. Their results, with ours, seem to indicate that researchers interested in studying critical appraisal skills should not substitute the easily estimated variable of journal article reading habits for the more difficult to determine variable of critical appraisal skill.

The critiques we received from both the experts and the novices tended to focus on the limitations of the study. Some respondents made introductory comments such as "Overall, this was a good article, but. . ." and then proceeded to list many specific limitations. Because of the very general nature of these comments, we chose not to incorporate them into our content analysis. Conclusions about the clinical usefulness of the study in light of the identified weaknesses were absent from almost all of the experts' and the novices' critiques. The focus on the limitations and the lack of clinical conclusions may be related to the way in which we presented the research task to potential subjects. We asked subjects to "critique" the study, assuming that the term "critique" implies an assessment of the strengths, weaknesses, and clinical applicability of the findings. The subjects, however, apparently interpreted "critique" narrowly as identification of limitations within the study. In future studies, researchers may wish to use the term "critical appraisal" rather than "critique," and provide a definition for respondents.

An obvious limitation of this study is the small response rate of 26.7%. This type of response rate has, however, been seen in another study related to development of critical appraisal skills of medical students.[8] In that study, only 5% of the medical students who were offered $25 to complete a 4- to 6-hour computer-aided instruction program designed to improve critical appraisal skills actually completed the project. Our study required the subjects to develop a free-form critique, which required much more time than, for example, completion of a standardized questionnaire. We assumed that our response rate would be low, but decided that the spontaneity of responses that would be obtained from the free-form critique was worth the low response rate. In addition, the subjects were all graduates of the same entry-level physical therapy program, limiting the generalizability of the results. Because of the time commitment required to complete the free-form critique, we chose to sample individuals who might participate based on their sense of commitment to their alma mater. Because the most frequently cited reason for participating in the study was a commitment to the university, we believe this limitation to generalizability was justified.

Two other limitations are related to the small sample and subgroup sizes. First, the boundaries among subgroups were not always based on theoretical distinctions, but were sometimes determined in ways that equalized subgroup sizes. This procedure may have obscured some group differences by

including individuals with intermediate levels of a variable with those with high or low levels. Second, the identification of trends was based on researcher-defined differences, rather than statistically determined differences; other researchers may have been either more conservative or more liberal in identifying important differences among various subgroups.

Further research on this topic should attempt to have a larger sample size, from different programs, to allow statistical comparison on important variables and greater generalizability across therapists. It would also have been interesting to have participants note whether they consulted reference materials in the preparation of the free-form critiques. In addition, it might be wise to permit respondents a choice of articles to allow them to demonstrate their critical appraisal skills in the context of familiar content. Another area for future research is the process of critical appraisal, that is, identifying critical steps that experts and novices take in analyzing and drawing conclusions from research literature. Examination of whether groups differ in their ability to discern the most important strengths and weaknesses in a study would also be useful.

Summary

Experts and novices differed most in the identification of construct validity concerns and were most similar in the identification of internal validity concerns. Only one factor—content-specific expertise—was associated with differences in critiquing ability among the novices.

References

1 Sackett DL, Haynes RB, Tugwell P. Cited by: Bennett KJ, Sackett DL, Haynes RB, et al. A controlled trial of teaching critical appraisal of the clinical literature to medical students. *JAMA.* 1987;257:2451–2454.

2 Commission on Accreditation in Physical Therapy Education. *Evaluative Criteria for Accreditation of Education Programs for the Preparation of Physical Therapists.* Alexandria, Va: American Physical Therapy Association; 1990.

3 Rothstein JM. Editor's note: Caveat emptor and conference abstracts. *Phys Ther.* 1990;70: 277–278.

4 *1990 Active Membership Profile Report.* Alexandria, Va: American Physical Therapy Association; 1991.

5 Linzer M, DeLong ER, Hupart KH. A comparison of two formats for teaching critical reading skills in a medical journal club. *J Med Educ.* 1987;62:690–692.

6 Linzer M, Brown JT, Frazier LM, et al. Impact of a medical journal club on house-staff reading habits, knowledge, and critical appraisal skills. *JAMA.* 1988;260:2537–2541.

7 Gehlbach SH, Bobula JA, Dickinson JC. Teaching residents to read the medical literature. *J Med Educ.* 1980;55:362–365.

8 Ashber JR. Letter to the editor. *JAMA.* 1990; 264:3143.

9 Bennett KJ, Sackett DL, Haynes RB, et al. A controlled trial of teaching critical appraisal of the clinical literature to medical students. *JAMA.* 1987;257:2451–2454.

10 Fitzgerald FT. From Galen to Xerox: the authoritarian reference in medicine. *Ann Intern Med.* 1982;96:245–246.

11 Blumenfield S. Gerontology journal club: a continuing education modality for experienced social workers in an acute care hospital. *Gerontologist.* 1985;25:11–14.

12 Casey M, Shafer P. One method of developing head nurses: the journal club. *Nursing Management.* 1982;13(4):36–37.

13 Draper V. Electromyographic biofeedback and recovery of quadriceps femoris muscle function following anterior cruciate ligament reconstruction. *Phys Ther.* 1990;70:11–17.

14 Cook T, Campbell D. *Quasi-experimentation: Design and Analysis Issues for Field Settings.* Chicago, Ill: Rand McNally & Co; 1979: chap 2.