Case Specificity of Standardized Patient Examinations: Constency of Performance on Components of Clinical Competence Within and Between Cases

Colliver, Jerry A.; Markwell, Stephen J.; Vu, Nu Viet; Barrows, Howard S.

How to cite

COLLIVER, Jerry A. et al. Case Specificity of Standardized Patient Examinations: Constency of Performance on Components of Clinical Competence Within and Between Cases. In: Evaluation & the health professions, 1990, vol. 13, n° 2, p. 252–261. doi: 10.1177/016327879001300208

This publication URL: https://archive-ouverte.unige.ch/unige:26113
Publication DOI: 10.1177/016327879001300208

Case Specificity of Standardized-Patient Examinations : Consistency of Performance on Components of Clinical Competence Within and Between Cases

Jerry A. Colliver, Stephen J. Markwell, Nu Viet Vu and Howard S. Barrows

Eval Health Prof 1990 13: 252
DOI: 10.1177/016327879001300208

CASE SPECIFICITY OF STANDARDIZED-PATIENT EXAMINATIONS

Consistency of Performance on Components of Clinical Competence Within and Between Cases

JERRY A. COLLIVER
STEPHEN J. MARKWELL
NU VIET VU
HOWARD S. BARROWS
Southern Illinois University

The psychometric properties of nine separate components of clinical competence (e.g., data collection, diagnosis, and management) assessed with standardized-patient cases were studied by Colliver et al. (1989). The findings of that study suggested the hypothesis that performance on a given component of clinical competence (e.g., data collection) would show less consistency when measured on different cases than would performance on different components (e.g., data collection, diagnosis, and management) measured within the same case. Thus, the purpose of the present study was to test this hypothesis. In general, the results supported the hypothesis. Specifically, generalizability coefficients were generally smaller for a given component across cases than for a given case across components. The recommendations implied by these results are: (a) More objective scoring methods are needed, and (b) The evaluation procedures for the components of clinical competence should be constructed as each case evaluation procedure is being constructed according to a clear conceptualization of the components which is developed in advance for all cases.

EVALUATION & THE HEALTH PROFESSIONS, Vol. 13 No. 2, June 1990 252-261

A Post-Clerkship examination, based on standardized-patient cases, has been used to assess clinical competence of senior medical students in four classes at Southern Illinois University (SIU) School of Medicine. Students are assessed on their total examination performance, as well as on nine separate components of clinical competence. The components assessed are data collection, as assessed by checklist completed by the standardized patient, data collection as assessed by written responses provided by students, working hypothesis, data interpretation, test selection, test interpretation, diagnosis, management, and working knowledge.

Using data from each of the four classes tested to date, Colliver, Vu, Markwell, and Verhulst (1989) evaluated the psychometric properties of these separate components of clinical competence. The results showed that the reliabilities of the components of clinical competence as they have been measured to date in the Post-Clerkship examination were generally low. Nearly two-thirds of the generalizability coefficients (inter-case reliabilities based on k = 10 cases) were less than .30. The low reliabilities of the components of clinical competence showed case specificity. Case specificity refers to the commonly reported finding that performance on one medical case (whether actual, simulated, or paper-and-pencil) is a poor predictor of performance on another case (Elstein, Shulman, & Sprafka, 1978). In addition, the pattern of intercorrelations among the components suggested the hypothesis that the case specificity was such that performance on a given component (e.g., data collection) would show less consistency when measured on different cases than would performance on different components (e.g., data collection, diagnosis, and management) within the same case.

The purpose of the present study was to test this hypothesis. To do so, a generalizability analysis was performed for each case, to determine the consistency of performance across the different components of clinical competence measured in that case. These results were compared with those of the generalizability analyses, performed by Colliver et al. (1989), that assessed the reliabilities of the separate components. Each of these generalizability analyses had been run for a given component of clinical competence, to determine the consistency of performance on that component across different cases.

METHODS

CLINICAL COMPETENCE EXAMINATION

Examinations of clinical competence were given to all senior medical students in four successive classes at Southern Illinois University School of Medicine. The examinations were administered to about 70 students in each of the four classes, about half-way through their last year of medical school. The 1986 examination contained 13 live, standardized patient cases; the 1987 exam, 17 such cases; the 1988 exam, 17 cases; and the 1989 exam, 18 cases. Details of the development, administration, and scoring of the clinical competence examination are presented elsewhere (Williams et al., 1987). In brief, the examination cases, all based on actual cases, were selected from a list of chief complaints and diagnoses of patient cases which faculty expect graduating students to be able to evaluate and in some cases manage. The dimensions of clinical competence to be assessed were determined and instruments were developed for collecting necessary data about student performance.

Students proceeded through the examination by first seeing each patient, performing a focused history and physical and, on occasion, providing the patient with tentative diagnostic conclusions, management plans, and patient education. Patients recorded the actions performed on the history and physical examination and their perceptions of the student's interpersonal skills and clinical service. Students then answered a series of short answer questions designed to elicit findings pertinent to the problem, their working hypothesis, their plans for laboratory investigation, their interpretation of laboratory findings, their diagnosis and management plan. Short-answer responses were evaluated using a preset answer key with predetermined weights assigned to each question. Standards of acceptable performance for each case were proposed by the case author and reviewed by a committee using current standards of quality care for the medical problem, and the expected level of performance of a graduating student.

TABLE 1
Generalizability Coefficients for Class of 1986

Agreement Across Cases		*Agreement Within Cases*	
Components of Clinical Competence	*Generalizability Coefficient*	*Case*	*Generalizability Coefficient*
Data collection – checklist (10)	.56	1 (6)	.49
		2 (6)	.75
Data collection – written (2)	.68	3 (5)	.67
		4 (5)	.36
Working hypothesis (7)	.57	5 (2)	.76
		6 (4)	.67
Data interpretation (2)	.68	7 (8)	.63
		8 (6)	.23
Test selection (7)	.57	9 (3)	.51
		10 (5)	.83
Test interpretation (5)	.48	11 (5)	.66
Diagnosis (8)	.45		
Management (10)	.40		
Working knowledge (6)	.35		
Mean	.53	Mean	.60

NOTES: Table contains generalizability coefficients showing, for each component of clinical competence, agreement across cases, and, for each case, agreement of different components within that case. All coefficients are based on k = 10 cases or components, projected with a D study approach. The number of cases or components on which G study analysis was performed is in parentheses.

COMPONENTS OF CLINICAL COMPETENCE

A number of components of clinical competence were tested on each case, although not all components were tested on every case. Nine of the components were tested on most cases and thus were used for analysis in the present study. The nine components studied were data collection (checklist completed by standardized patient), data collection (written responses of students), working hypothesis, data interpretation, test selection, test interpretation, diagnosis, management, and working knowledge. The numbers of cases that measured each component of clinical competence are given in parentheses in Tables 1 through 4. For example, for the class of 1986 shown in Table 1, data

TABLE 2
Generalizability Coefficients for Class of 1987

Agreement Across Cases		*Agreement Within Cases*	
Components of Clinical Competence	*Generalizability Coefficient*	*Case*	*Generalizability Coefficient*
Data collection—checklist (13)	.50	1 (6)	.43
		2 (6)	.51
Data collection—written (7)	.50	3 (6)	.42
		4 (6)	.37
Working hypothesis (9)	.30	5 (6)	.18
		6 (5)	.17
Data interpretation (7)	.28	7 (6)	.38
		8 (6)	.46
Test selection (10)	.04	9 (5)	.25
		10 (4)	.73
Test interpretation (6)	.07	11 (5)	.70
		12 (3)	.51
Diagnosis (13)	.06	13 (4)	.33
		14 (10)	.37
Management (20)	.06	15 (9)	.38
		16 (6)	.55
Working knowledge (11)	.43	17 (3)	.71
Mean	.25	Mean	.44

See Note for Table 1.

collection—checklist was measured on 10 cases, data collection—written was measured on 2 cases, and working hypothesis was measured on 7 cases. Similarly, the numbers of components of clinical competence measured by each case are given in parentheses in the four tables. For example, for the class of 1986 in Table 1, 6 components were measured on case 1, 6 on case 2, and 5 on case 3.

STATISTICAL METHODS

Generalizability analyses were performed to assess the two types of agreement evaluated in the study (Cronbach, Glesser, Nanda, & Rajaratnam, 1972). First, a generalizability analysis was performed for each component of clinical competence. These analyses provided information about the agreement of scores obtained on a given component as measured on different cases (i.e., inter-case reliability). These results were available from the Colliver et al. (1989) study.

TABLE 3
Generalizability Coefficients for Class of 1988

Agreement Across Cases		*Agreement Within Cases*	
Components of Clinical Competence	*Generalizability Coefficient*	*Case*	*Generalizability Coefficient*
Data collection – checklist (14)	.59	1 (4)	.40
		2 (6)	.28
Data collection – written (7)	.27	3 (6)	.44
		4 (5)	.56
Working hypothesis (5)	.00	5 (7)	.30
		6 (6)	.26
Data interpretation (7)	.00	7 (7)	.34
		8 (7)	.29
Test selection (10)	.02	9 (5)	.69
		10 (5)	.31
Test interpretation (5)	.50	11 (5)	.73
		12 (3)	.00
Diagnosis (10)	.00	13 (4)	.56
		14 (6)	.00
Management (17)	.20	15 (3)	.78
		16 (6)	.29
Working knowledge (10)	.32		
Mean	.21	Mean	.40

See Note for Table 1.

Second, a generalizability analysis was performed for each case. These analyses provided information about the agreement of the scores obtained on the different components within a given case. Naturally, a G-study approach to obtain variances components was the initial step in the computation of all generalizability coefficients. Thus, different G-study analyses were based on different numbers of data points (i.e., clinical competence components or cases). The numbers of data points on which these initial G-study analyses were performed are given in parentheses in Tables 1 through 4. For example, for the class of 1986, the G-study analysis for data collection – checklist was performed on 10 data points (cases); for data collection – written, 2 cases; and for working hypothesis, 7 cases. Similarly, for case 1, the G-study analysis was performed on 6 data points (components of clinical competence); for case 2, 6 components; and for case 3, 5 components. A D-study analysis was then used to combine the variance components to obtain generalizability coefficients subjected to a

TABLE 4
Generalizability Coefficients for Class of 1989

Agreement Across Cases		*Agreement Within Cases*	
Components of Clinical Competence	*Generalizability Coefficient*	*Case*	*Generalizability Coefficient*
Data collection – checklist		1 (4)	.42
		2 (7)	.48
Data collection – written (6)	.37	3 (5)	.47
		4 (5)	.77
Working hypothesis (5)	.19	5 (6)	.51
		6 (3)	.77
Data interpretation (6)	.11	7 (4)	.34
		8 (5)	.45
Test selection (9)	.32	9 (3)	.56
		10 (4)	.66
Test interpretation (5)	.09	11 (2)	.39
		12 (2)	.87
Diagnosis (13)	.21	13 (5)	.68
		14 (7)	.62
Management (12)	.24	15 (2)	.00
		16 (4)	.19
Working knowledge (10)	.40		
Mean	.24	Mean	.51

See Note for Table 1.

Spearman-Brown adjustment so that all results would be based on a common or standard number of data points (components or cases), thus ensuring the comparability of all generalizability coefficients reported in the article. This standard was arbitrarily chosen to be k = 10 data points. Thus, all generalizability coefficients reported were based on a standard number of k = 10 components or cases using a D study approach. That is, for the clinical competence components, the coefficients were based on k = 10 cases; for the cases, the coefficients were based on k = 10 components.

RESULTS

In general, the generalizability coefficients obtained for each of the components of clinical competence were smaller than the generalizability coefficients obtained for each of the cases. (See Tables 1,

2, 3, and 4.) For example, for the class of 1989, the generalizability coefficients for the components ranged from .09 to .40 with a mean of .24, whereas the coefficients for the cases ranged from .00 to .87 with a mean of .51. (See Table 4.) Similar results were obtained for the classes of 1986, 1987, and 1988. (See Tables 1, 2, and 3.) The averages of the generalizability coefficients for the components for these classes were .53, .25, and .21, whereas the averages for the cases were .60, .44, and .40, respectively.

DISCUSSION

The results of the analysis support the prediction that, on average, performance on a given component would show less agreement across different cases than would performance on different components within the same case. The most parsimonious explanation of these findings would seem to be that they are an effect of a given grader scoring all components of clinical competence within a case and different graders scoring the same component on different cases. In other words, there appears to be a kind of "halo effect" at work in the scoring of different components within a case. For example, detailed performance on one component within a case may cause the grader to give the student the benefit of the doubt when scoring a more general response on another component in that case. The implication of this explanation is that more objective scoring methods are needed, especially for the written components of the examination, which tend to have the lowest reliabilities.

Another plausible explanation of the findings stems from the method used to construct the evaluation procedures for the cases and the separate components of clinical competence. The evaluation procedure includes the checklist items and short answer questions with scoring keys, used to evaluate student performance. Each case and its evaluation procedure is constructed as a whole by a given faculty member, and different cases are constructed by different faculty members. After a case is constructed, the case author identifies the items that best correspond to the separate components of clinical competence, thus defining the components of clinical competence for

that case. This construction of the evaluation procedure for a case as a whole with the identification of components after the fact could easily account for the greater generalizability of different components within a case as compared to that of the same component measured on different cases. The recommendation implied by this explanation of the findings is that components of clinical competence should assume a more central role in the construction of the evaluation procedure for the cases. To accomplish this, a clear conceptualization of the components of clinical competence should be developed and communicated to the case authors before the cases are constructed. The case authors, then, should construct the components as the case is being constructed; that is, the construction of the evaluation procedure for the components should be an integral part of the construction of the case evaluation procedure, rather than an afterthought. This approach should result in components that have a more consistent meaning across cases and thus greater inter-case reliability than has been reported.

A more general explanation of such findings has been simply "case specificity," in the sense that any given medical case is seen as a unique whole (Elstein, Shulman, & Sprafka, 1978; Norman, 1982). In this sense, then, the different parts of a case necessarily "hang together," and different cases necessarily differ. The implication of this explanation is that any attempt to measure the separate components of clinical competence is necessarily futile, because clinical performance is unique to a case or at least to the content measured by a case.

Although the latter explanation can account for the findings of the present study, the first two explanations considered here are more parsimonious and provide testable recommendations for more reliable assessment of the separate components of clinical competence. In summary, these recommendations are (a) to develop more objective scoring methods for the examination and (b) to construct the evaluation procedures for the components of clinical competence as the case evaluation procedure is being constructed according to a clear conceptualization of the components developed in advance for all cases. It would seem that these two recommendations should be tested before accepting the more general "case specificity" explanation.

REFERENCES

Colliver, J. A., Vu, N. V., Markwell. S. J., & Verhulst, S. J. (in press). Psychometric Properties of clinical competencies assessed with standardized patient cases. In I. Hart & R. Harden (Eds.), *Developments in Assessing Clinical Competence*. Montreal: Heal Publications.

Cronbach, L. J., Glesser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Norman, G. R. (1982, September 4). Medical problem solving and the illusion of content specificity. *Professions Education Researcher Notes*, pp. 10-12.

Williams, R. G., Barrows, H. S., Vu, N. V., Verhulst, S. J., Colliver, J. A., Marcy, M., & Steward, D. (1987). Direct, standardized assessment of clinical competence. *Medical Education*, *21*, 482-489.