Predictive Validity of Two Medical Specialty Preference Inventories

MARK L. SAVICKAS
JOAN S. BRIZZI
LORI A. BRISBIN
LURA L. PETHTEL

Practitioner-based scales containing items related to medical practice more accurately predicted students' medical specialty choice than did student-based scales containing generic interest items.

A physician's specialty choice has important personal, economic, and societal consequences (Czinkota & Johnston, 1983). Therefore, medical school faculty members and health care planners encourage medical students to make personally suitable and socially viable choices. Many students seek counseling help in making this choice because they must choose a specialty before they have sufficient experience and information (Savickas, Alexander, Osipow, & Wolf, 1985). Efforts to provide such help have included several attempts to develop medical specialty preference scales. Two different types of specialty preference scales are currently available. The authors of the Medical Specialty Preference Scales (Gough, 1979) and the Medical Specialty Preference Inventory (Zimny & Senturia, 1976), however, have reported only preliminary validity data for their scales. Our study addressed this deficiency by examining the predictive validity of the Medical Specialty Preference Scales (MSPS) and the Medical Specialty Preference Inventory (MSPI).

Although the authors of the MSPS and MSPI used different scale construction strategies, both strategies evolved from the pioneering research of Strong and Tucker (1952) who devised SVIB medical specialty interest scales for surgeons, internists, pathologists, and psychiatrists. Strong and Tucker administered these scales to 783 medical students in 1950. Unfortunately, the majority of these students scored highest on the psychiatry scale, and 10 years later the four scales did not differentiate the specialist groups (Tucker & Strong, 1962). Campbell (1966) reevaluated the 783 SVIB blanks and developed new specialty scales using more stringent criteria. This reevaluation also produced disappointing results. Campbell concluded that measurable specialty interests may not appear until after physicians have practiced a specialty for several years because their interests may be modified by their work experiences. This conclusion led to different strategies in constructing the MSPS (Gough, 1979) and MSPI (Zimny & Senturia, 1976).

Athelstan and Paul (1971) reasoned that even if the interests of a mature specialist do not crystallize until after several years

Mark L. Savickas is associate professor, and Joan S. Brizzi is a research associate in the Behavioral Sciences Department, Northeastern Ohio Universities College of Medicine, Rootstown, Ohio. Lori A. Brisbin is a psychology assistant at the Child and Adolescent Center in Canton, Ohio, and a doctoral candidate in the Counseling Psychology Program, Kent State University, Kent State, Ohio. Lura L. Pethtel is assistant director of the Office of Medical Education, Northeastern Ohio Universities College of Medicine.

of practice, the characteristics that eventually lead to a specialty may be measurable in students. Therefore, they abandoned the traditional approach of basing scales on practitioners. Instead, they used student interests that differentiated specialty position 11 years later to construct six student-based scales. The surgery, obstetrics-gynecology, psychiatry, and pediatrics scales correctly identified 56% of the specialists in the developmental sample. Despite this promising beginning, the Athelstan and Paul scales were never cross-validated and they became obsolete when the 1966 revision of the SVIB did not include many items in these scales. Subsequently, Gough (1979) followed Athelstan and Paul's rationale and strategy to devise student-based medical specialty scales for the Strong-Campbell Interest Inventory (SCII). Gough's MSPS provided excellent differentiation in the developmental sample. The scales have not been cross-validated, however. In fact, we were unable to find any published studies dealing with the MSPS other than Gough's original report.

Zimny used a different strategy to construct the MSPI. He used the traditional strategy of basing keys on practitioners, but he did not use generic interest items like those in the SCII. Instead, Zimny used items directly related to the practice of medicine. The MSPI scoring system deals with the job functions of each specialty because interests may co-vary more with the functional dimensions within medicine and less with the formal divisions between specialties. The one predictive validity study we found for the MSPI reported an overall hit rate of 51% for 385 students from five medical schools. The hit rates for students at each of the five schools were 47%, 47%, 48%, 55%, and 65% (Zimny, 1980, p. 418). Our study examined the predictive validity of both the MSPS (Gough, 1979) and the MSPI (Zimny & Senturia, 1976). This report may also serve as a useful model of one approach to evaluating predictive validity of counseling measures.

METHOD

Measures

The MSPS uses the items in the Strong-Campbell Interest Inventory (SCII). The SCII items deal with school subjects, occupational titles, and types of people. To devise the MSPS, Gough administered an earlier version of the SCII to 956 freshman medical students. After these students had begun medical practice, an analysis of their responses identified items that differentiated among students who had chosen different specialties. Gough selected 40 items to constitute each of 10 student-based scales: medicine (MED), obstetrics/gynecology (OBG), pediatrics (PED), psychiatry (PSY), surgery (SUR), family medicine (FAM), anesthesiology (ANES), ear-nose-throat (ENT), pathology (PATH), and radiology (RAD).

The MSPI (Zimny & Senturia, 1976) consists of 199 items about medical activities and settings. Students rate each item on a 7-point scale. The student responses are scored on 40 factors that characterize the practice of medicine. A student's scores on these 40 factors are compared to the scores obtained by reference groups of practicing physicians. This comparison yields scores of six practitioner-based scales: internal medicine (MED), obstetrics-gynecology (OBG), pediatrics (PED), psychiatry (PSY), surgery (SUR), and family medicine (FAM).

Data Collection and Analyses

Two consecutive classes of students at a state supported medical school in the midwest voluntarily participated in our study during their clerkship year. The class of 1984 ($N = 71$) completed the MSPS in January 1983; the class of 1985 ($N = 73$) completed the MSPI in January 1984. A com-

mercial service scored the scales and profiled them in standard score units for each student. From the profiles, we predicted which specialty each student would enter. The predicted specialty was the one with the highest standard score. All students who took the MSPI except one had at least one score above 70, which was the minimum cutoff score required to make a prediction in Zimny's study (1980). Predictions were made only for those students whose specialties could be predicted by the 10 scales of the MSPS or the six scales of the MSPI, so that for the MSPS, *N* equaled 68, and for the MSPI, *N* equaled 64. Nineteen months after the students responded to the inventories, the Dean for Student Affairs reported the specialty that each student actually entered.

Zimny (1980) discussed the validity of first year residency choice as an initial criterion, and indicated that specialty choice remained consistent through the second year of residency for 76% of the students in his sample. To discover how many students in our samples changed specialty choice and how these changes affected predictive validity, we gathered information regarding changes after students completed their first year of residency. This waiting period also yielded the residency choices of students who had selected a transitional first year residency with no declared choice of specialty.

To test the predictive validity of the inventories, each student's predicted specialty was compared to the specialty that the student entered. This allowed us to see whether the predicted and actual choice agreed. Two different statistics, an overall hit rate and Cohen's (1960) kappa, provided information about agreement between the predicted and actual choices for each set of scales. Hit rate equals the proportion of students who actually entered the specialty that received the highest score on the profile. Computing the hit rate permitted direct comparison of the hit rates in this study to those in Zimny's study. Gough did not use hit rate in validating his scale.

Although hit rate provides an important index of the usefulness of a test, it should not be interpreted independently of other types of information such as the proportion of students who were predicted not to choose a specialty but did choose it, or the proportion of accurate predictions that could be expected by chance. For example, Zimny reported a "51% overall level of predictive accuracy compared to a chance expected agreement of 17%" (Zimny, 1980, p. 418). He computed chance expectancy as 1 out of 6, or 17%, because the MSPI has six scales. This value ignores the marginal distributions that are used in chi-square derivatives, like kappa, to determine chance expectancy. In fact, Zimny even pointed out later in his report that a reason why the internal medicine scale had the highest hit rate may have been because more students selected that specialty, thereby increasing the likelihood that a prediction of internal medicine would be accurate. A single solution to the problems of reporting only proportion of agreement and inaccurately calculating chance expectancy is found in Cohen's (1960) coefficient kappa.

The maximum value of kappa is 1.0, and occurs when there is perfect agreement between predicted and actual outcomes. When obtained agreement equals expected chance agreement, kappa equals 0. Landis and Koch (1977) suggested the following interpretations of kappa: kappa less than .40 represents poor agreement beyond chance; kappa between .40 and .75 represents fair to good agreement beyond chance; and kappa larger than .75 represents excellent agreement beyond chance.

In addition to overall hit rates and overall kappas for the two inventories, hit rates and kappas were calculated individually

for each scale on each inventory. These calculations have less stability than overall kappa because of the small number of students in each specialty. The computation of a kappa for an individual specialty involved collapsing all specialties into one category except the one of interest. For example, a kappa for Gough's MED scale involved constructing a 2×2 table with predicted MED or non-MED for the rows, and actual choice of MED or non-MED for the columns. Calculation of kappa for each scale involved repeating this same process. The overall kappa amounts to a complex weighted average of these individual kappas.

RESULTS

Medical Specialty Preference Scales

Of the 68 MSPS (Gough, 1979) profiles used in the analyses, only 13 resulted in accurate predictions. These results yielded a 19% hit rate and an overall kappa of .15. Table 1 reports the hit rate and kappa for each of the 10 MSPS scales. Only three of these scales had good hit rates: medicine (67%), surgery (100%), and family medicine (50%). None of the 10 scales had a kappa above .34. Thus, even the three scales with good hit rates had poor kappas. For example, the surgery scale had a 100% hit rate and a kappa of .21. Therefore, even though the hit rate was perfect, the proportion predicted to enter surgery was less than the proportion that actually entered. Specifically, 14 students entered surgery residencies but only two of these were predicted to do so. The 100% hit rate means that of the two students who were predicted to enter surgery, both did. This leaves 12 students who entered surgery but were predicted to enter some other specialty. Thus, the surgery scale had a high hit rate but did not identify an adequate proportion of the students who entered surgery residencies.

Five students changed their choice of specialty after the first year of residency. Of these five, one student had been predicted to enter psychiatry and had originally selected psychiatry, but subsequently changed to family medicine. Another student had been predicted to enter radiology, entered family medicine, and changed to radiology. The other three students' scores resulted in inaccurate predictions that remained inaccurate despite change of specialty choice. Changes did not appreciably affect hit rate or kappa values.

One student entered a transitional first year residency without declaring specialty choice until the end of that first year. At the end of the first year of residency, the student, who had been predicted by the MSPS to enter psychiatry, selected obstetrics. The score for psychiatry was 55, however, and the profile showed little differentiation. Addition of this student to the results did not appreciably affect either the kappa value or the hit rate.

Excluding the one student who did not declare specialty choice until the second year, 65 of 70 students (93%) remained consistent in choice of specialty through the second year. This figure is considerably higher than Zimny's (1980) reported result of 76%.

Medical Specialty Preference Inventory

Of the 64 MSPI (Zimny & Senturia, 1976) profiles used in the analyses, 38 resulted in accurate predictions. These results yielded a 59% hit rate and a kappa of .48. Table 1 reports the hit rate and the kappa for each of the six MSPI scales. The surgery and family medicine scales both had hit rates and kappas over .60. The medicine scale had an excellent hit rate but a poor kappa. The remaining three scales had both poor hit rates and kappas. As one would expect, the three scales with the higher proportion of students enter-

TABLE 1

Hit Rates and Kappa Coefficients for 10 MSPS and 6 MSPI Scales

	MSPS ($n=68$) (Gough, 1979)			MSPI ($n=64$) (Zimny & Senturia, 1976)		
Scale	Proportion of Students Selecting Each Specialty	Hit Rate	Kappa	Proportion of Students Selecting Each Specialty	Hit Rate	Kappa
Medicine (MED)	.34	.67	.09	.31	.70	.32
Obstetrics/Gynecology (OBG)	.07	.25	.23	.06	.33	.31
Pediatrics (PED)	.07	.33	.32	.08	.33	.26
Psychiatry (PSY)	.03	.06	.06	.05	.33	.37
Surgery (SUR)	.21	1.00	.21	.28	.64	.62
Family Medicine (MED)	.16	.50	.34	.22	.65	.61
Anesthesiology (ANES)	.03	.00	−.03			
Ear-Nose-Throat (ENT)	.04	.00	−.04			
Pathology (PATH)	.02	.06	.06			
Radiology (RAD)	.03	.00	−.05			

ing that specialty had the highest hit rates. The kappas that correct for distribution of students across the specialties, however, indicted that only two of these three scales had good agreement beyond chance. In fact, the psychiatry scale had a lower hit rate but a higher kappa ($k = .37$) than the medicine scale ($k = .32$). This occurred because the psychiatry scale more effectively predicted students who would not choose the specialty. The psychiatry scale's negative hit rate was .97 compared to the medicine scale's negative hit rate of .76, although this may be due more to the small number of students entering psychiatry rather than the scale itself.

Seven students from the class of 1985 changed specialty choice after their first year of residency. Of these seven students, one had been predicted to enter psychiatry but initially selected family medicine. This student's subsequent choice was psychiatry. Three students' choices had been accurately predicted according to their first year choices (one in family medicine, one in surgery, and one in medicine) but these resulted in inaccurate predictions given choices following the first year. Two of these students (who had originally chosen family medicine and medicine) entered emergency medicine, a specialty not considered by the MSPI. The student who had initially selected surgery transferred into family medicine. The three remaining students who changed specialty choice were originally considered to be inaccurately predicted and remained so. These changes did not significantly affect hit rate and kappa values.

Four students from the class of 1985 entered transitional residencies for 1 year without stating subsequent specialty choice. Of these four students, three selected specialties that could not be predicted by the MSPI and one student was predicted to enter family practice and subsequently did. Again, the changes did not appreciably affect the hit rate and kappa values. Excluding the four transitional students from the total 1985 sample ($N = 73$), 62 of 69 students (90%) remained constant in their specialty choice. This figure also is considerably higher than Zimny's 76%.

DISCUSSION

The predictive validity of the MSPS and MSPI differed markedly. The MSPS (Gough, 1979) displayed poor predictive validity. The MSPS's overall hit rate of 19% and kappa of .15 indicted that the MSPS yielded accurate predictions little beyond what one would expect simply by chance. Although the rationale for student-based scales is appealing, these scales have not yet been successfully cross-validated. At this time, we must require data supporting predictive validity of the MSPS before using it as a tool in individual career counseling with medical students.

The MSPI (Zimny & Senturia, 1976) fared much better. The overall hit rate of 59% fell at the high end of the range reported by Zimny (1980). This figure favorably compares to the conclusion drawn by Holland, Magoon, and Spokane (1981, p. 283) that "in a six category system, most inventories have a hit rate of about 40% plus or minus 5%." The 59% hit rate was especially impressive because the MSPI predicted among overlapping specialties within one occupation rather than among distinct occupational categories. The kappa of .48 fell in the fair range, but seems quite strong compared to the kappas that Rothman (1985) reported for agreement between expressed and actual specialty choice.

The data for the MSPS and MSPI came from different groups of students. Thus, we must be cautious in comparing the relative efficacy of the two sets of scales. Until we have a study that uses both the MSPS and the MSPI with one group of students, however, we may hypothesize that the predictive validity of the MSPI significantly exceeds that of the MSPS. In fact, our data

indicated that the MSPI predicted accurately three times more than the MSPS. Without a great deal of further research, we can only speculate as to whether this difference in predictive validity arises from the difference in scale type (student-based versus practitioner-based) or item type (generic versus occupation-specific).

In addition to its better predictive validity, the MSPI has some distinct advantages for counseling. The MSPI can be used to teach students about medicine's occupational functions and specialty structure. A counselor can discuss what the student likes to do based on the student's scores for the 40 functions, and then discuss what the student may like to be based on the student's scores for the six specialties. By comparing what a student wants to do and wants to be, a counselor can address specialty choice indecision caused by the functional overlap between specialties and the personality differences within specialties. In summary, based on its predictive validity and its teaching possibilities, the MSPI seems more useful than the MSPS for counseling medical students about specialty choice.

REFERENCES

Athelstan, G.T., & Paul, G.J. (1971). New approach to the prediction of medical specialization: Student-based Strong Vocational Interest Blank scales. *Journal of Applied Psychology, 55,* 80–86.

Campbell, D.P. (1966). *Reanalyses of Strong's interest data from medical specialists* (Final report for Contract No. OE-6-10-148, Project No. 5-8400). Washington, DC: U.S. Office of Education.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Czinkota, M.R., & Johnston, W.J. (1983). Choosing a career and specialty: When do students decide. *Health Care Management Review, 8,* 43–51.

Gough, H.G. (1979). *Gough medical specialty preference scales: A report for counselors.* Palo Alto, CA: Consulting Psychologists Press.

Holland, J.L., Magoon, T.M., & Spokane, A.R. (1981). Counseling psychology: Career interventions, research, and theory. *Annual Review of Psychology, 32,* 279–305.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 22,* 159–174.

Rothman, A.I. (1985). Statements on career intentions as predictors of career choices. *Journal of Medical Education, 60,* 511–516.

Savickas, M.L., Alexander, D.E., Osipow, S.H, & Wolf, F.M. (1985). Measuring specialty indecision among career-decided students. *Journal of Vocational Behavior, 27,* 356–367.

Strong, E.K., & Tucker, A.C. (1952). The use of vocational interest scales in planning a medical career. *Psychological Monographs, 66,* (Whole No. 341).

Tucker, A.C., & Strong, E.K. (1962). Ten-year follow-up of vocational interest scores of 1950 medical college seniors. *Journal of Applied Psychology, 46,* 81–86.

Zimny, G.H. (1980). Predictive validity of the Medical Specialty Preference Inventory. *Medical Education, 14,* 414–418.

Zimny, G.H., & Senturia, A.G. (1976). *Medical Specialty Preference Inventory.* St. Louis: Department of Psychiatry, St. Louis University Medical Center.