

Predicting Success in Academic Achievement of Major Seminarians in Papua New Guinea: A Comparison of Cognitive Test Results and Grade Point Averages

David G. Songy

Abstract

Cognitive tests were administered to seminarians in Papua New Guinea ($N = 135$) and compared to Grade Point Average (GPA) scores collected 3 months later to discover the predictive validity of such tests. Multiple regression analyses and a logistic regression procedure indicated significant predictive value for certain subtests of the Culture Fair Intelligence Test (CFIT) and the Michigan Test of English Language Proficiency (MTELP) to GPA. While GPA means varied across different classes of seminarians, the test score means were more consistent. The tests were found to be reliable on measures of internal consistency and could predict low achievement with 76% accuracy. Further exploration of the use of cognitive tests related to predicting achievement in this cultural setting is warranted.

Key Words: Seminarian, Assessment, Papua New Guinea, CFIT, MTELP, Roman Catholic

Introduction

Selecting candidates for a specialized field can be a complex process (Anastasi & Urbina, 1997). The decision to accept or reject an individual is based on knowledge of that person's abilities in relationship to a unique role in society. The importance of the selection process is especially evident in the Roman Catholic Church that seeks out men with a certain level of education, dedication to ministry, spiritual maturity, and an ability to be faithful to vows of celibacy and obedience. The universal Code of Canon Law states:

Can. 241 - ss1. The diocesan bishop is to admit to the major seminary only those who are judged capable of dedicating themselves permanently to the sacred ministries in light of their human, moral, spiritual and intellectual characteristics, their physical and psychological health and their proper motivation. (Canon Law Society of America [CLSA], 1983, p. 83).

In more recent years, the shift in Third World Catholic Churches from relying on missionary priests to ordaining their own indigenous clergy has made necessary the introduction of selection standards adapted to the local cultures (CLSA, 1983).

Can. 242 - ss1. Each nation should have a program for priestly formation which is to be determined by the conference of bishops in light of the norms issued by the supreme authority of the church and which is also to be approved by the Holy See . . . this program is to define the main principles for imparting formation in the seminary as well as general norms which have been adapted to the pastoral needs of each region or province [italics added]. (CLSA, 1983, p. 83).

Such adaptations will enable seminaries in the Third World to select candidates who are able to complete a program of priestly formation that will ensure a high quality of ministry to the local Church.

Papua New Guinea context

In the case of selecting seminarians in Papua New Guinea, an increasing number of candidates have entered the seminary in the past 30 years, whose initial numbers were 29 in 1964 and grew to 218 by 1994 (Aerts, 1994). Two new major seminaries were built in the early 1990s to complement Holy Spirit Seminary, the only major seminary in the country since the early 1960s.

Despite the high numbers entering the major seminary, less than 20% of those entering reach ordination. One can only speculate as to the reasons for the large numbers leaving in the middle of their training. Some choose another walk of life. Others cannot handle the academic challenges and after expending several years of effort are asked to leave. Of those who complete the training process and are ordained, some find that they are not adequately prepared for the life and ministry of a priest, while others flourish.

Seminary professors and rectors try to establish standards of achievement that will contribute to the success of seminarians, but the task is complex. Papua New Guinea is the most culturally diverse country in the world (Whiteman, 1984), with 700 languages spoken by a population of four million. Although English is the official language of education, few speak it outside a school setting; and the quality of English varies according to geographical regions (Richardson, 1987). The Catholic Church is young in Papua New Guinea, and the depth of faith and knowledge of religion varies widely among seminarians. Given the formidable task of evaluating potential success in selecting seminary candidates at various stages of formation, I will only address one important aspect of the process, the prediction of academic achievement. If it were possible to predict academic success or failure more accurately, limited resources could be used more efficiently and that group of young men who struggle with academic performance could be guided to other vocational choices. While the educational system in Papua New Guinea is English based, the quality of education varies from region to region and English proficiency varies greatly (Richardson, 1987). Therefore, it would be helpful to determine the English reading abilities of students. Greater knowledge of overall intellectual functioning could help determine remediation or treatment alternatives for those who display certain weaknesses but have the potential of

doing better. Furthermore, in the process of examining academic achievement, seminary rectors and professors could determine reasonable standards of expectation for academic performance.

Anastasi and Urbina (1997) have described the importance of criterion-prediction validation procedures which “indicate the effectiveness of a test in predicting an individual’s performance in specified activities” (p. 118). This is especially relevant when using tests in screening personnel for specialized fields or entrance into educational or training institutions (Anastasi & Urbina, 1997). The Graduate Record Examination (GRE), for example, is a test of aptitude given to potential graduate students because it has been found to be a good predictor of graduate degree completion (House & Johnson, 1993). The actual criterion used to measure graduate school completion is the degree issued, but the GRE has provided a simple, quick method of predicting to this criterion. Establishing the validity of predicting academic achievement through the use of objective tests in Papua New Guinea would contribute to the screening process currently in use.

Several studies have been conducted concerning intelligence testing in Papua New Guinea (Ord, 1971), though these have decreased in the past 20 years due to the lessening numbers of psychologists in the country (Richardson, 1987). Ord (1971) provided a history of the use of intelligence tests in the country, largely with illiterate populations. Doubts have been expressed about the suitability of psychometric tests in diverse populations, because the validity of such tests has been established only with a very specific population (Helms, 1992; Sibaya, Hlongwane, & Makunga, 1996). Such tests may be biased, especially when used across cultural contexts. Language is often a confounding factor (Van den Berg, 1992), and little research has been conducted using standardized instruments with people of various cultures (Sibaya, Hlongwane, & Makunga, 1996). Reliability, validity, and norms need to be established whenever using an instrument outside the cultural context in which it was created (Anastasi & Urbina, 1997). Recognizing these important concerns, some researchers have developed instruments that they have found to be reliable and valid across many cultural settings, including the Michigan Test of English Language Proficiency (MTELP; Division of Testing and Certification, English Language Institute [ELI], 1977) and the Culture Fair Intelligence Test (CFIT; Institute for Personality and Ability Testing [IPAT], 1973).

Without valid and reliable screening instruments, professors and rectors will continue to rely on previous academic performance as the sole objective measure in the selection of seminary candidates. My objective in this study was to ascertain whether additional instruments would aid the screening process. Specifically, my objective was to study the reliability and validity of the MTELP and CFIT in predicting academic success for seminarians in Papua New Guinea.

Method

Participants

Participants were recruited from two major seminaries in Papua New Guinea: Good Shepherd Seminary in Banz, Western Highlands Province ($n = 91$) and Holy Spirit Seminary in Bomana, National Capital District ($n = 48$). Of the original 139 volunteers, three were dropped from the study because of incomplete data. After initial review of the data, one subject was dropped as an outlier, having scored more than three standard deviations below others on certain tests. The final number of participants included in this study were 135 (Good Shepherd Seminary, $n = 87$; Holy Spirit Seminary, $n = 48$) with a median age of 23 years ($M = 24.04$, $SD = 3.17$).

All participants had completed Grade 12 before entering the seminary. The participants identified 66 distinct languages as their first language spoken and represented 16 of the country's 18 provinces. Many were from the coastal regions (42%), while a large number came from the highlands (39%). The islands regions were also represented (25%), though their numbers were fewer since participants were not solicited from the major seminary in East New Britain. Fifty participants were first year students; 47 were in their second year; and 38 were from the third year. While Holy Spirit Seminary has a six year training program, only seminarians from the first three years participated in the study, having a similar academic program as the seminarians from Good Shepherd Seminary. There were no major differences in age or cultural background between participants from Holy Spirit Seminary and those from Good Shepherd Seminary. Both seminaries recruit members from all provinces in Papua New Guinea.

Materials and procedures

Two instruments were used in this study: the MTELP (ELI, 1977) and the CFIT (IPAT, 1973). Both tests were chosen because they could be administered in a group setting and had been used in various cross-cultural settings previously (ELI, 1977; IPAT, 1973). First, all of the tests were administered in large group settings after participants had completed the process for giving informed consent. Then their instructors were asked two questions about each participant: (a) What do you predict will be the grade this student receives for your course at the end of this term? (b) In your estimation, if this student were to perform to the best of his potential in your course, what would you predict his grade should be at the end of this term? They were instructed to give a percentage grade, according to the grading policy for major seminaries in Papua New Guinea: A: 100-90%; B: 89-80%; C+: 79-70%; C: 69-60%; D: 59-50%; F: <50%. Finally, three months after the original testing of participants, grade reports were collected from the seminaries.

Culture Fair Intelligence Test.

The CFIT, Scales 2 and 3, is a timed test that has eight sections and takes a total of 25 minutes to complete. The eight sections cover four areas: Series, Classifications, Matrices, and Conditions (Topology). Each item stem has four symbols. The test taker selects from several options the symbol that belongs with the other four. This test was developed to provide an instrument that

would minimize the influences of cultural learning and still offer a reliable and valid estimation of "g", that is, a measure of general intelligence (IPAT, 1973). In other words, this test, without relying on language or other culturally determined factors, measures the general factor referred to as intelligence. Cattell (1973) explained that the test directly measures fluid intelligence, that which is "free of particular investments" (p. 6) such as language or concrete information learned in a particular context. He argued that fluid intelligence is distinct from crystallized intelligence, but highly correlated. The CFIT has been used among various cultural groups in the United States, as well as in Germany, Pakistan, Canada, Finland, Mexico, and the United Kingdom (IPAT, 1973). Jordheim and Olsen (1963) used Scale 2 in Micronesia, finding similarities in intelligence levels across a wide variety of cultural groups.

Cronbach's alpha for the 96-item scale was .85. The subscale reliability coefficients were acceptable: Series, .72; Classification, .61; Matrices, .76; and Conditions, .70.

Michigan Test of English Language Proficiency

The MTELP was introduced in the early 1960's as part of a battery to estimate whether a student who spoke English as a Second Language (ESL) could perform well in an English language university and what level and number of courses that student could undertake given a certain proficiency level in English. The complete test battery includes a 30 minute written composition on a specific topic and a test of aural comprehension, both of which offer a more comprehensive view of the individual's capacity to perform in an English language university. However, I chose to omit these sections and administered only the three subtests that were able to be conducted in a group setting: Grammar, Vocabulary, and Reading Comprehension. The entire test of 100 multiple choice items has a 75 minute time limit. In the Grammar section, each item is in the form of a dialogue between two people. One part of the dialogue has been omitted, and the test taker is given several options to choose from. The Vocabulary section has two types of questions: One type requires substituting the given word for a synonym. The second type requires filling in a word from several options according to the context of the sentence. The Comprehension section has four reading passages ranging from 100 to 350 words, and each passage is followed by five multiple-choice questions.

Reliability and validity studies have involved people of various cultural groups: Romance languages, Chinese, Japanese, Germanic, and Arabic (ELI, 1977). Over a three-year period from 1978 to 1980, Baldauf and Dawson (1980) compared MTELP scores with grade point averages (GPA) obtained from students at teacher training colleges in Papua New Guinea and found the MTELP was a reliable and valid instrument to predict academic performance.

Cronbach's alpha for the 100-item scale was .84. The subscale reliability coefficients were acceptable: Grammar, .63; Vocabulary, .69; and Comprehension, .80.

Grade Point Average

The GPA of each student was obtained. In addition to this, since all of the seminary instructors had taught the seminarians for two previous terms and had knowledge regarding their academic potential, they were asked to provide ratings of predicted GPAs three months before the conclusion of the study. Stone (1994) has conducted research showing that teacher's ratings may be compared with group ability tests results as alternative predictors of ability.

Results

I first examined the descriptive statistics for the CFIT and MTELP as well as actual GPA scores of each seminary class. All subtests were normally distributed except the Comprehension subtest of the MTELP, which was positively skewed. A multivariate analysis of variance (MANOVA) revealed significant differences in scores among the six classes of seminarians, Wilk's Lambda (5, 129) = .62, $p < .001$. The post hoc test of the GPA, CFIT, and MTELP indicated that this was due to differences in actual GPA scores only. I divided the participants into three groups based on actual GPA scores: low achievers (<-1 SD), average achievers (-1 SD to 1 SD), and high achievers (>1 SD). The resultant groups had 21, 96, and 18 members. Three groups similar in size were then formed by taking the low and high achievers and a group of 20 randomly selected average achievers. Analyses could then be conducted on the total number of participants ($N = 135$) as well as the three groups that were balanced according to size and achievement ($n = 59$).

Table 1: Scores for the Culture Fair Intelligence Test Subtests, means, standard deviations, range and maximum possible

	Mean	S.D.	Range	Maximum
Scale 2				
Series	6.90	2.57	0-12	12
Classification	6.54	1.85	0-11	14
Matrices	6.16	2.95	0-11	12
Conditions	3.41	1.90	0-8	8
Scale 3				
Series	5.26	1.72	0-10	13
Classification	4.39	2.09	0-10	14
Matrices	3.27	1.68	0-7	13
Conditions	4.39	1.82	0-8	10
Total	40.31	9.69	16-63	96

Table 2. Scores for the Michigan test of English language proficiency subtests, means, standard deviations, range and maximum possible

	Mean	S.D.	Range	Maximum
Grammar	27.65	4.25	15-37	40
Vocabulary	21.17	4.71	10-35	40
Comprehension	6.59	4.07	0-18	20
Total	55.41	10.34	30-84	100

Table 1 shows the descriptive statistics of performance on the CFIT, and Table 2 indicates those for the MTELP. On the Comprehension subtest ($\bar{M} = 6.59$, $SD = 4.07$), the skewed distribution of scores is likely due to the fact that this was the last subtest administered, and many participants did not complete any of this subtest due to time constraints.

Table 3. Means, Standard Deviations and Range of Grade Point Averages according to Seminary classes

	Mean	S.D.	Range	n
<i>Good Shepherd Seminary</i>				
Year 1	71.47	6.24	59.60-87.00	40
Year 2	62.44	9.14	49.25-83.25	25
Year 3	71.49	6.36	55.83-83.00	22
Subtotal	68.88	8.23	49.25-87.00	87
<i>Holy Spirit Seminary</i>				
Year 1	69.64	5.28	60.33-77.83	9
Year 2	77.14	5.02	69.17-88.67	22
Year 3	73.19	6.84	61.25-87.40	17
Subtotal	74.33	6.35	60.33-88.67	48
TOTAL	70.82	8.03	49.25-88.67	135

Table 3 shows the actual GPA among various classes of seminarians. It is important to note the large differences in means among the groups (range = 14.70).

Table 4. Means and Standard Deviations for CFIT and MTELP according to Seminary classes

	CFIT Mean	CFIT. SD	MTELP Mean	MTELP SD	<u>n</u>
Holy Spirit Seminary					
Year 1	34.67	13.11	50.44	5.53	9
Year 2	41.00	10.48	59.68	9.53	22
Year 3	40.18	11.27	58.41	12.27	17
Good Shepherd Seminary					
Year 1	40.78	9.56	54.03	10.31	40
Year 2	41.64	8.33	54.04	8.33	25
Year 3	39.68	7.82	54.95	12.04	22
Total	40.31	9.69	55.41	10.34	135

CFIT = Culture Fair Intelligence Test. MTELP = Michigan Test of English Language Proficiency

Table 4 lists the descriptive statistics of the CFIT and MTELP according to various classes of seminarians. In the case of the CFIT, the means are more similar among groups (range = 6.97). The same is true with the MTELP (range = 9.24). It should be noted that the first year class at Holy Spirit Seminary (n = 9) scored the lowest mean for both the CFIT and MTELP, while the lowest actual GPA mean was given to the second year class at Good Shepherd Seminary (n = 25).

The correlation matrix for the CFIT, MTELP, and actual GPA is presented in Table 5a. It is interesting to note the low correlation between actual GPA and the tests.

Table 5a. Pearson Correlation Matrix of Grade Point Average, CFIT, and MTELP Scores

	GPA	CFIT						MTELP						
		Scale 2	Scale 3			Total			Vocab.	Gram	Compr	Total		
	Series	Classif.	Matr.	Cond.	Series	Classif.	Matr.	Cond.	Total	Vocab.	Gram	Compr	Total	
GPA	1	.037	.163*	.085	.038	-.004	.133	.062	.258**	.162	.325***	.240**	.338***	.380***
CFIT														
Scale 2														
Series	.037	1	.303***	.407***	.232**	.360***	.042	.312***	.187*	.654***	-.019	.098	.005	.013
Classif.	.163*	.303***	1	.194*	.077	.223**	.155*	.196*	.103	.472***	.115	.193*	.122	.180*
Matr.	.085	.407***	.194*	1	.229**	.399***	.196*	.391***	.206**	.714***	.089	.238**	.125	.188*
Cond.	.038	.232**	.077	.229**	1	.118	.205**	.214**	.358***	.512***	.155	0.135	.087	.160
Scale 3														
Series	-.004	.360***	.223**	.399***	.118	1	.121	.424***	.331***	.622***	-.058	.227**	-.004	.065
Classif.	.133	.042	.155*	.196*	.205**	.121	1	.204**	.210**	.453***	.218*	.249**	.243**	.298***
Matr.	.062	.312***	.196*	.391***	.214**	.424***	.204**	1	.283***	.627***	.158	.154	.117	.181*
Cond.	.258**	.187*	.103	.206**	.358***	.331***	.210**	.283***	1	.543***	.186*	.295***	.236**	.299***
Total	.162	.654***	.472***	.714***	.512***	.622***	.453***	.627***	.543***	1	.173*	.338***	.182*	.290***
MTELP														
Vocab.	.325***	-.019	.114	.089	.155*	-.058	.218**	.158*	.186*	.173*	1	.434***	.512***	.836***
Gram.	.240**	.098	.193*	.238**	.135	.227**	.249**	.154*	.295***	.338***	.434***	1	.377***	.757***
Compr.	.338***	-.046	.122	.124	.087	-.004	.243**	.117	.236**	.182*	.512***	.377***	1	.782***
Total	.380***	.013	.180*	.188*	.160	.065	.298***	.181*	.299***	.290***	.836***	.757***	.782***	1

GPA = Grade Point Average. CFIT = Culture Fair Intelligence Test. MTELP = Michigan Test of English Language Proficiency; Classif. = Classification. Matr. = Matrices. Cond. = Conditions. Vocab. = Vocabulary. Gram. = Grammar. Compr. = Comprehension.; *p < .05. **p < .01. ***p < .001

Table 5b. Pearson Correlation Matrix of Grade Point Average, CFIT, and MTELP Scores

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. GPA	--													
2. Series (2)	.04	--												
3. Classif. (2)	.16	.30**	--											
4. Matr. (2)	.09	.41**	.19	--										
5. Cond. (2)	.04	.23*	.08	.23*	--									
6. Series (3)	.00	.36**	.22*	.40**	.12	--								
7. Classif. (3)	.13	.04	.16	.20	.21*	.12	--							
8. Matr. (3)	.06	.31**	.20	.39**	.21*	.42**	.20*	--						
9. Cond. (3)	.26*	.19	.10	.21*	.36**	.33**	.21*	.28**	--					
10. Total CFIT	.16	.65**	.47**	.71**	.51**	.62**	.45**	.63**	.54**	--				
11. Vocab.	.33**	-.02	.11	.09	.16	-.06	.22*	.16	.19	.17	--			
12. Gram.	.24*	.10	.19	.24*	.14	.23*	.25*	.15	.30**	.34**	.43**	--		
13. Compr.	.34**	-.05	.12	.12	.09	.00	.24*	.12	.24*	.18	.51**	.38**	--	
14. Total MTELP	.38**	.01	.18	.19	.16	.07	.30**	.18	.30**	.29**	.84**	.76**	.78**	--

GPA = Grade Point Average. Series (2) = Culture Fair Intelligence Test (CFIT), Scale 2, Series subscale. Classif. (2) = CFIT, Scale 2, Classification subscale. Matr. (2) = CFIT, Scale 2, Matrices subscale. Cond. (2) = CFIT Scale 2, Conditions subscale. Series (3) = CFIT, Scale 3, Series subscale. Classif. (3) = CFIT, Scale 3, Classification subscale. Matr. (3) = CFIT, Scale 3, Matrices subscale. Cond. (3) = CFIT, Scale 3, Conditions. Total CFIT = CFIT composite score. Vocab. = Michigan Test of English Language Proficiency (MTELP) Vocabulary. Gram. = MTELP Grammar. Compr. =

*MTELP Comprehension. Total MTELP = MTELP composite score. * $p < .01$. ** $p < .001$*

Multiple regression procedures were performed using the CFIT and MTELP with the dependent variable alternately being the actual GPA of all participants, teachers' predicted GPA, and teachers' predicted best GPA. Using a multiple regression technique with the GPA as the dependent variable and CFIT and MTELP as independent variables, I found that composite scores on both the CFIT and MTELP were significant predictors of GPA, $F(2, 132) = 11.4$, $p < .0001$. Exploring the teachers' predicted GPA as the dependent variable, I found that composite scores on both the CFIT and MTELP were also significant predictors of predicted GPA, $F(2, 132) = 17.7$, $p < .0001$. The model that best fit the data was between teachers' predicted best GPA and the CFIT and MTELP. The multiple regression procedure indicated that the MTELP and CFIT were significant predictors of teachers' predicted best GPA scores, $F(2, 132) = 18.2$, $p < .0001$. I then used a multiple regression procedure, breaking the MTELP into three subscales and found that only the Comprehension subtest of the MTELP was significant in a model to predict the teachers' predicted GPA and teachers' predicted best GPA. Breaking the CFIT into its eight subscales, only the CFIT Scale 3 Conditions Subtest was significant in a model to predict actual GPA.

In order to estimate the unique contribution of each subtest in predicting GPA, I calculated the squared semipartial correlations. Those subtests most highly correlated with GPA were CFIT Scale 3 Conditions Subtest ($sr^2 = .21$), MTELP Comprehension Subtest ($sr^2 = .16$), MTELP Vocabulary Subtest ($sr^2 = .15$), and CFIT Scale 2 Classification Subtest ($sr^2 = .12$). Using this procedure, the three groups that were balanced according to size and achievement ($n = 59$) produced higher correlations: CFIT Scale 2 Classification Subtest ($sr^2 = .35$), MTELP Vocabulary Subtest ($sr^2 = .27$), CFIT Scale 2 Matrices Subtest ($sr^2 = .23$), and MTELP Comprehension Subtest ($sr^2 = .21$).

Table 6. Logistic regression using CFIT and MTELP to predict low achievers (-1S.D. GPA) within equal size groups

	Predicted	Predicted	Percent
	Group 1	Group 2	Correct
Observed Group 1	16	5	76%
Observed Group 2	4	34	89%
Overall Correct			85%

CFIT = Culture Fair Intelligence Test. MTELP = Michigan Test of English Language Proficiency. GPA = Grade Point Average. Group 1 = ($\leq -1S.D.$). Group 2 = ($> -1S.D.$)

A logistic regression procedure was then used to determine how well the CFIT and MTELP could predict actual GPA given groups that were balanced according to size and achievement ($n = 59$). Table 6 indicates 76% accuracy in classifying low achievers in Group 1 (≤ -1 SD actual GPA) and an overall 85% accuracy in distinguishing low achievers in Group 1 from average and

high achievers in Group 2 ($> -1SD$ actual GPA), $(11, n = 59) = 27.7, p < .01$. The most significant variables in the equation were CFIT Scale 2 Matrices Subtest and CFIT Scale 2 Classification Subtest, while the only MTELP subtest that approached significance was the Vocabulary Subtest. This procedure was less accurate in predicting low achievement when only the MTELP scores were used. While overall accuracy was 81%, the accuracy for predicting low achievers was only 57%, and none of the subtests were statistically significant.

Discussion

The data analysis suggests that the MTELP and CFIT are useful in predicting academic achievement of seminarians in Papua New Guinea. Both instruments were found to be reliable according to Cronbach's measure of internal consistency. While there were stronger correlations between the MTELP and GPA, suggesting that the MTELP is a good measure of English abilities relevant to achievement in this culture, the CFIT stood out as more accurate in distinguishing low achievers. The CFIT seemed to be a better indicator of general intelligence. However, claims of internal validity can only be made cautiously, given the wide range among the means of the various classes of seminarians. As others have noted in using GPA as a dependent variable (Morrison & Morrison, 1995), it is not always the best indicator of achievement. Oldfield and Hutchison (1997) have argued that an objective measure of achievement would provide a more reliable dependent variable than GPA, which may be confounded by the limited range in GPA and varying standards among instructors and programs. The results of this present study tend to support their argument.

Those subtests most useful in distinguishing low achievers employed constructs related to understanding matrices, classification, and vocabulary. The Matrices and Classification subtests of the CFIT require more complex thinking processes than the Series or Classification subtests. So while there was no significant difference between students identifying the next object in a series or classifying an object, the low achievers had difficulty with the complex tasks of filling in a matrix design or identifying similar conditions among figures. Low achievers also found the Vocabulary portion of the MTELP more difficult than the Grammar section, possibly because filling in a correct grammatical answer is easier than guessing the correct meaning of a new word. Further evaluation of the differences in performance on subtests, including the wise use of time, and their relationship to academic achievement in Papua New Guinea may prove useful in helping low achievers.

At this point, there is not enough evidence to recommend that the MTELP and CFIT or any of their subtests be used as the sole criteria for predicting academic success of seminarians. While my study generated 76% accuracy statistically in predicting low achievement, a 24% margin of error would not be acceptable. It would be best to combine MTELP and CFIT test results with other measures, which could include high school GPA, a written essay test, an individual interview to determine English communication skills, and letters of

recommendation that answered specific questions regarding the academic potential of a student. A decision for admission to the academic program of the seminary could be based on this larger pool of information.

My study was conducted with a specific population of young men studying in a seminary setting, so the results may not be generalized to other groups in Papua New Guinea without further study. Other limitations in my study include the use of GPA as the single dependent variable, the fact that all tests were administered in groups, and that all of the tests were timed. The individual abilities of seminarians were not fully investigated given some of these limitations.

Overall, the evidence in my study supports the notion that cognitive measures can help predict academic performance in Papua New Guinea seminaries. It raises the question of how to measure achievement in this setting accurately and highlights those subtests which may point to areas of further study. Further research might include switching the order in which subtests are given to control for the influence of time, or administering tests with no time limit (IPAT, 1973). Finally, the development of more accurate measures of achievement to serve as criterion variables with which to contrast cognitive test results would be desirable. This would seem to be essential to the screening and evaluation process necessary for the successful formation of priests in Papua New Guinea.

References

- Aerts, T. 1994, Towards localizing the Catholic priesthood, in T. Aerts (ed.), *The Assembly Hall at Bomana* (pp. 8-30), Holy Spirit Seminary, Bomana, Papua New Guinea.
- Anastasi, A. & Urbina, S. 1997, *Psychological testing* (7th edn.), Prentice Hall, Upper Saddle River, NJ.
- Baldauf, R. B. & Dawson, R. L. 1980, The predictive validity of the Michigan Test of English Language Proficiency for teacher trainees in Papua New Guinea, *Educational and Psychological Measurement*, 40, 1201-1205.
- Canon Law Society of America [CLSA] 1983, *Code of Canon Law: Latin-English edition*, Canon Law Society of America, Washington, DC.
- Cattell, R. B. 1973, The nature and measurement of intelligence, in Institute for Personality and Ability Testing [IPAT], *Technical Supplement for the Culture Fair Intelligence Tests Scales 2 and 3* (pp. 5-8), IPAT, Champaign, IL.
- Division of Testing and Certification, English Language Institute [ELI] 1977, *Michigan test of English language proficiency*, Ann Arbor: University of Michigan.
- Helms, J. E. 1992, Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, 47, 1083-1101.
- House, J. D., & Johnson, J. J. 1993, Graduate record examination scores and academic background variables as predictors of graduate degree completion, *Educational and Psychological Measurement*, 53, 551-556.

- IPAT 1973, *Technical Supplement for the Culture Fair Intelligence Tests Scales 2 and 3*, Institute for Personality and Ability Testing, Champaign, IL.
- Jordheim, G. D., & Olsen, I. A. 1963, The use of a non-verbal test of intelligence in the trust territory of the Pacific, *American Anthropologist*, 65, 5.
- Morrison, T., & Morrison, M. 1995, A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the graduate record examination with graduate grade point average representing the criterion of graduate success, *Educational and Psychological Measurement*, 55, 309-316.
- Oldfield, K., & Hutchison, J. R. 1997, Predictive validity of the graduate record examination with and without range restraints, *Psychological Reports*, 81, 211-220.
- Ord, I. G. 1971, Assessing the cognitive capacities of non-literate New Guinea adults, Paper prepared for the Conference on 'Cultural Factors in Mental Test Development, Application, and Interpretation', Istanbul, Turkey.
- Richardson, B. 1987, Psychology in Papua New Guinea: A brief overview, in G. H. Bowers & A. M. Turtle (eds), *Psychology Moving East: The Status of Western Psychology in Asia and Oceania* (pp. 289-303), Westview Press, Boulder.
- Sibaya, P. T., Hlongwane, M., & Makunga, N. 1996, Giftedness and intelligence assessment in a third world country: Constraints and alternatives, *Gifted Education International*, 11, 107-113.
- Stone, B. J. 1994, Group ability tests versus teachers' ratings for predicting achievement, *Psychological Reports*, 75, 1487-1490.
- Van den Berg, A. R. 1992, *Position Paper on Test Construction and Use in South Africa*, Human Sciences Research Council, Pretoria.
- Whiteman, D. L. 1984, Melanesia: Its people and cultures, in D. L. Whiteman (ed.), *An Introduction to Melanesian Cultures* (pp. 85-104), Melanesian Institute, Goroka, Papua New Guinea.

Author

Fr David Songy is a Capuchin Franciscan priest psychologist and Director of Counseling Services at the Pontifical North American College in Rome. A former missionary to the Southern Highlands of Papua New Guinea, he returned to PNG to conduct research on cognitive testing and achievement as part of his doctoral dissertation. Recently an Associate Professor of Pastoral Theology at St. John Vianney Theological Seminary in Denver, Colorado (USA), Father Songy continues to specialize in the formation of seminarians and assessment of priesthood candidates.

