**Creating Equality in Testing Student
with English as a Second Language**

Jean Streifling

Kim is a Korean girl enrolled in the 6th grade of an English-medium classroom. Although English is not her first language, she communicates well with her classmates and she performs above the class average in her daily school activities and assignments. Kim's written performance, displaying comprehension, and understanding on written tests and examinations, are problematic, because it is significantly below the class average.

Several possible reasons for Kim's poorer performance on written tests and examinations could be suggested.

- First, she possibly suffers from examination anxiety causing her to be unable to think clearly and remember what she has learned.

- Second, it may be that she is receiving a lot of help with her daily work assignments at home, or from other students. If this were the case, Kim could be completing all the assignments without really understanding what she is doing and therefore might be unable to answer questions on examinations.

- A third possible reason for the difference between her daily work performance and her examination performance could be Kim's limited understanding of written English. If this were the case, she might be compensating for her limited ability to decode written English in her daily assignments by asking help from others or looking at pictures and other contextual clues, helps which are unavailable in the setting of a written examination.

This third possible reason for the difference between her daily work and her written examination performance has initiated the following research.

1. Is there a significant difference between "English as a Second Language" (ESL) students' rankings of their multiple intelligences on a written test in English, and their ranking of the same intelligences on a pictorial test?

2. When considered by test type, is there a significant difference between "English as a Second Language" (ESL) students' rankings of their multiple intelligences, and "English as a First

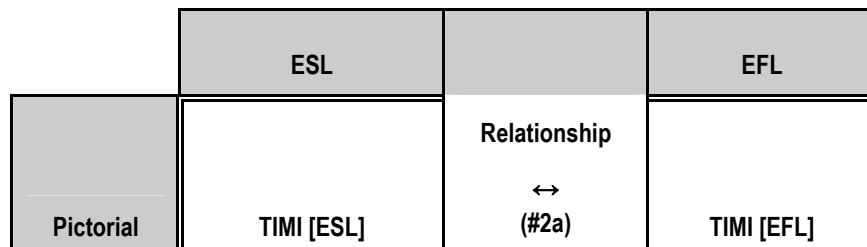Language" (EFL) students' rankings of their multiple intelligences?

3. Is there a significant difference between "English as a First language" (EFL) students' rankings of their multiple intelligences on a written test in English, and their ranking of the same intelligence on a pictorial test?
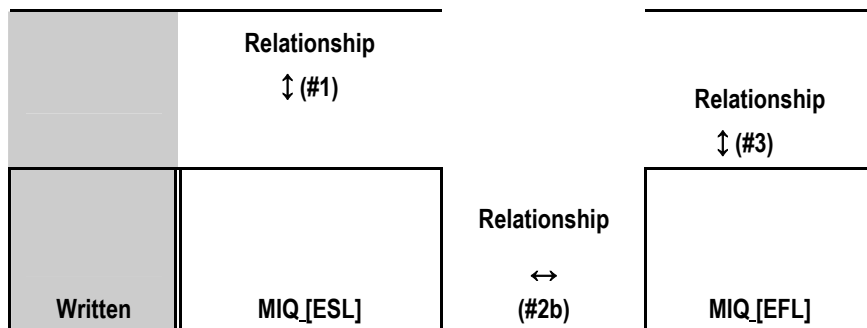
If the underlying reason for Kim's problem was her limited understanding of written English, a problem shared by other ESL students, then it would be expected that a significant difference would be found between results on parallel written and pictorial tests given to ESL students. It was expected that there would not be significant differences between the accumulated group results for the two language groups on the written or pictorial tests since preferences across multiple intelligences are probably similarly distributed across each language group. When the two types of tests were administered to EFL students, it was expected that the results would be similar, being strongly positively correlated.

Conceptual Framework

A conceptual framework, (Figure 1) was developed to assist visualization of the three different relationships being explored. The first relationship (#1) depicted was between the pictorial and the written assessment test for the ESL students. The second relationship (#2) represented was between the two language groups for both types of tests. The third relationship (#3) considered was between the pictorial and the written assessment test for the EFL students.

Figure 1

	ESL		EFL
Pictorial	**TIMI [ESL]**	**Relationship** ↔ **(#2a)**	**TIMI [EFL]**

	Relationship ↕ (#1)		Relationship ↕ (#3)
Written	**MIQ_[ESL]**	**Relationship ↔ (#2b)**	**MIQ_[EFL]**

Testing ESL Students

A number of earlier studies (Kifer 1994, McKeon 1994, O'Malley and Waggoner 1984, Kaufman (1996) appear to be related to the scope of this study. The findings of Kifer (1994), are very interesting. He found that a major failing of assessment tests is that they measure ability under crippling circumstances, at least for some people. In taking an assessment test, an ESL student has many disabilities that a native speaker does not have.

Kifer (1994) gives several good reasons why ESL students test unreliably. First, their reading speed is slow. Second, their vocabulary is tiny compared to that of a native English-speaking student. Kindergarten students in the U.S. understand far more words than ESL students can reasonably be expected to learn in a few years. Third, the ESL student's vocabulary is nonstandard; they don't know aspects of our language that English speakers take for granted. Fourth, the English words lose a lot of meaning in the translation process. Fifth, the ESL student comes from a different culture: even when they understand the words, they might misunderstand the question. In addition, a person learns best when he or she is prepared to receive the information, but these students are having to tackle strange questions in a strange language about strange situations. To some extent, the student's past education is a hindrance since it has been teaching him or her how to cope with a "different world." It could be said that usually ESL students don't fail assessment tests, but the assessment tests fail the students.

How can these tests fail the student? How can a valid and reliable test be rendered a worthless assessment tool? The very act of creating a valid and reliable test may exclude significant factors, such as understanding, that are much more important to the student's total success than the factors which the test is measuring.
 In his experience of teaching ESL students for almost nine years, Kifer (1994) has observed time and again the relationship between poor English ability and the unachieved potential for academic success among ESL students. It is clear that

learning in a language that you speak almost automatically is a different proposition to learning about information in a language that you have yet to master.

There is another compounding factor for ESL students that McKeon (1994) describes. In the different countries of origin, the curricular sequences, content objectives, and instructional methodologies may differ dramatically from American educational practices. This too must affect the student's level of comprehension in assessment tests.

The issue of assessment of ESL students is of great concern to educators. It is the standardized written test in particular that is a source of debate because it is on the basis of test scores on written tests that an ESL student is often mis-assigned to lower curriculum tracks or special education (LaCele-Peterson and Rivera 1994).

Kifer (1994) conjectures that, "No written assessment test provides the ESL student the opportunity to utilize all of his resources, and therefore, written assessment tests may not accurately predict academic ability."

In an article focusing on student achievement and the policy for bilingual education, the U.S. Department of Education (1996) reports that nearly 30 percent of ESL students attend school without any curriculum adapted to their linguistic requirements.

Researchers had expressed concern more than a decade before, about the lack of an adapted curriculum provided for ESL students. O'Malley and Waggoner (1984) provide a possible reason for such neglect in adapting the curriculum in the U.S. to the needs of the ESL students. Approximately half of all American teachers teach an ESL student at some time in their career, but there is no required course work focusing on strategies for teaching ESL students that a student teacher takes in their teacher training education.

The dramatic growth in linguistically diverse populations in the U.S.A. has significantly altered the population of American schools, and caught many administrators and teachers unprepared. It is imperative in an English-speaking school that the teachers of ESL students and the teachers of other subject areas work together for the effective education of language minority students (Kaufman 1996).

In this study the researcher poses a possible reason for the inequities between ESL and EFL student assessments. There needs to be a solution to the difficulty of evaluating ESL students' written tests, and teachers need to learn helpful strategies they can implement when teaching ESL students.

Method

In this study two alternative test strategies were compared, one a written assessment instrument, and the other, a pictorial instrument for assessing ESL students. The instruments chosen were both designed to evaluate students' preferences for utilizing the seven multiple intelligences identified by Howard Gardner (1983)—linguistic, logical-mathematical, musical, spatial, body-kinesthetic, intrapersonal, and interpersonal. Teele (1994) describes typical characteristics of each of the intelligences (Appendix #1). Teele postulated that the understanding and usage of these multiple intelligences should affect teaching methodologies and assessments. To help teachers easily analyze the multiple intelligences of their students, Teele (1992) created a pictorial instrument; the Teele Inventory of Multiple Intelligences (TIMI).

Instruments

One of the instruments used in this study was the TIMI. It has been utilized in a broad spectrum of applications since its inception, being administered to students at the preschool level through elementary, secondary, community college and institutions of higher education. Teele claims for this instrument test - retest studies' reliability (Teele 1992). It has been used in more than 450 different public and private schools throughout the United States, and in six other countries throughout the world. The inventory is designed to show that students possess different combinations of the seven intelligences, which gives evidence that they process information in many different ways.

The TIMI is a forced-choice pictorial inventory that contains 56 numbered pictures of panda bears representing characteristics of each of the seven intelligences and provides students twenty-eight opportunities to state a preference for one of two alternatives presented. The answer sheet is easily scored, and when tallied, the scores present a profile of the responses of each student enabling the teacher to identify the student's most dominant intelligences as indicated by the highest scores.

The second instrument used in this study was a written Multiple Intelligences Questionnaire (MIQ), designed by Batulayan (2001). This questionnaire was composed of 70 items, with each intelligence having an equal representation of 10 items each. Content validity was used to claim that the instrument logically measures an intended variable. The MIQ was subjected to reliability analysis and found to have a measured reliability coefficient of .85 Cronbach alpha.

The respondents were asked to rate the items in the MIQ on a three-point Likert scale designated by "disagree" as 1, "not sure" as 2, and "agree" as 3, so as to provide numeric data. Scores of 1-10 indicated a low level of dominant intelligence, 11- 20 indicated moderate or average, and 21-30 points meant a high

level of dominant intelligence. Fifty percent of the items were written as negative statements and the other 50 percent were written as positive statements. Any answers to a negative item in the questionnaire were recoded so that responses to these items could be appropriately combined with responses to positively stated items to form a scale score for each inelligence.

The hypothesis associated with the first research question was that a significant difference would be found between results on parallel written and pictorial tests for ESL students required to complete the written test in the English language. The TIMI together with the MIQ allows such comparison to be made.

Sample

The population used for this study were students enrolled in the Adventist International Institute of Advance Studies International (AIIAS) Elementary School. All the students enrolled in 4th grade through 8th grade, participated, giving a total population of 123 students. This was a combination of girls and boys from a variety of countries, mainly the Philippines, Korea, and seven other Asian countries who used English as their second language. The total number of ESL students was 111. The remaining 12 students came from homes where English was the first language spoken in the family environment.

Procedures

The researcher administered the assessment instruments to the students in grade four to eight classrooms at the AIIAS Elementary School. For the TIMI assessment test the students made a choice from pairs of activity pictures viewed on overhead transparencies. The students marked their choices on the answer sheet provided by the researcher. Approximately 30 minutes was taken to complete the test in each classroom. In the afternoon of the same day, the same students met together in an assembly room to answer the MIQ written questionnaire. Again, the students completed the test within 30 minutes. Help was provided so the students could understand directions for completing the test, but no help was given relating to the meaning of words in the questionnaire.

The students' responses on the TIMI, and the MIQ were scored by the researcher using the answer sheets developed for each test. The student's TIMI scores were calculated by tallying the sum of their choices for each item related to a specific intelligence. These scores were ranked from one to seven, with one being the intelligence most frequently indicated by the student. The scores from the MIQ were tallied and translated into ranks ranging from one to seven, with "1" representing the highest dominant intelligence.

The ranked data and the demographic information were analyzed using SPSS. For the first and third research questions, t-tests were used to compare the paired

samples. In this process the relationship between the results of the two instruments was investigated by calculating for the correlation coefficient between the two instruments. An independent group t-test was used to find an answer to the second research question, whether there was a significance difference between the results of the pictorial TIMI and the written MIQ for the two language groups.

Results

The results of the analysis of students' rankings using t-tests and correlations are described in the following sections.

ESL Rankings on Written vs. Pictorial Tests

The results from the ESL student's paired sample t-test revealed statistically significant differences between the TIMI and MIQ tests' mean ranks in all seven areas of intelligences (Table 1). However, the correlation coefficient "r" results for the TIMI and MIQ test (Table 1), indicated that the only statistically significant positive correlation between the two instruments was for math intelligence, (r = 0.273, p<0.01).

Table 1. **Paired Samples Correlation Test for ESL Students**

	t	df	Sig (2-tail)	Pearson "r"	p
Pair 1: TIMI linguistic - MIQ linguistic	-6.030	110	.000	-.029	.759
Pair 2: TIMI math - MIQ math	4.450	110	.000	.273	.004
Pair 3: TIMI spatial - MIQ spatial	-7.328	110	.000	.108	.259
Pair 4: TIMI music - MIQ music	6.239	110	.000	.145	.128
Pair 5: TIMI body/kinesthetic - MIQ body/kinesthetic	2.259	110	.026	-.036	.708
Pair 6: TIMI intrapersonal - MIQ intrapersonal	6.951	110	.000	.028	.769
Pair 7: TIMI interpersonal - MIQ interpersonal	-5.062	110	.000	-.058	.545

Rankings by ESL vs. EFL Student Groups

To investigate the differences between EFL speaker's and the ESL speaker's accumulated ranking scores in each intelligence, an Independent Group t-test was used. (Table 2).

The accumulated ranking scores showed significant differences between the two language groups in two multiple intelligences. On the written MIQ test, ESL and EFL students indicated a significantly different mean ranking of their music intelligence. The ESL students' intelligence was higher by two ranks (t = 3.326, df = 121, p = .001). Another difference was the mean ranking of linguistics. The ESL students showed a significant difference to the EFL students in the mean ranking of their linguistic intelligence on the pictorial TIMI. The ESL students' linguistic score was higher by 1.5 ranks (t = 2.987, df = 121, p = .003).

EFL Rankings on Written vs. Pictorial Tests

The results from the paired sample t-test comparing rankings on the TIMI and MIQ for the EFL students revealed no significant different rankings in five areas of intelligences (Table 3), but there were statistically significant differences between the TIMI and MIQ test's mean rank in spatial (t = 2.746, df = 11, p = 0,019) and interpersonal intelligences (t = 2.477, df = 11, p = 0.031). However, for the EFL students, the table of paired sample correlations between the TIMI and MIQ tests, only showed significant correlation in the music intelligence (r = 0.858, p<0,01) (Table 3).

Table 3. **Paired Samples Correlation Test for EFL Students**

	t	df	Sig (2-tail)	Pearson	p
Pair 1: TIMI linguistic - MIQ linguistic	1.117	11	.288	-.149	.644
Pair 2: TIMI math - MIQ math	.534	11	.604	.322	.307
Pair 3: TIMI spatial - MIQ spatial	-2.746	11	.019	.408	.188
Pair 4: TIMI music - MIQ music	-.453	11	.660	.858	.000
Pair 5: TIMI body/kinesthetic - MIQ body/kinesthetic	1.232	11	.244	.244	.444
Pair 6: TIMI intrapersonal - MIQ intrapersonal	1.789	11	.101	.094	.771
Pair 7: TIMI interpersonal - MIQ interpersonal	-2.477	11	.031	.207	.519

Discussion

In some ways the findings of this study were unexpected, puzzling and inconclusive. Perhaps it has raised more questions rather than providing answers. The first three of the following discussion sections parallel the research questions of the study.

The visual representation of the relationships in this study (Figure 1) has been used to report mean ranking of intelligences for each type of test, TIMI and MIQ, and both language groups of students, ESL and EFL (Figure 2). The first relationship, the vertical comparisons (#1) is between the pictorial and the written assessment test for the ESL students. The second relationship, represents the horizontal comparisons (#2a and 2b) between the two language groups for each type of tests. The third relationship, a vertical comparison (#3) is between the pictorial and the written assessment test for the EFL students.

***Figure 2.* Mean Ranks Profile of Multiple Intelligences**

	ESL			EFL	
	TIMI [ESL]			**TIMI [EFL]**	
	Interpersonal	2.8 (1st)		Interpersonal	2.3 (1st)
	Spatial	3.0 (2nd)		Spatial	3.0 (2nd)
	Linguistic	3.5 (3rd)		Body/Kinesthetic	3.3 (3rd)
Pictorial	Body/Kinesthetic	3.9 (4th)		Math	4.0 (4th)
	Math	4.2 (5th)	Relationship	Linguistic	5.0 (5th)
	Music	4.8 (6th)	↔	Music	5.2 (6th)
	Intrapersonal	5.8 (7th)	(#2a)	Intrapersonal	5.3 (7th)
	Relationship ↕ (#1)			Relationship ↕ (#3)	
	MIQ [ESL]			**MIQ [EFL]**	
	Math	3.3 (1st)		Body/Kinesthetic	2.6 (1st)
	Music	3.3 (1st)		Math	3.5 (2nd)
	Body/Kinesthetic	3.4 (3rd)		Interpersonal	3.9 (3rd)
Written	Interpersonal	4.1 (4th)		Intrapersonal	4.1 (4th)
	Intrapersonal	4.4 (5th)	Relationship	Linguistic	4.1 (4th)
	Spatial	4.6 (6th)	↔	Spatial	4.4 (6th)
	Linguistic	4.9 (7th)	(#2b)	Music	5.3(7th)

ESL Rankings on Written vs. Pictorial Tests

It had been hypothesized that a significant difference would be found between results on parallel written and pictorial tests for ESL students (Relationship #1-- Figure 2). The comparison of individual intelligence rankings between the two

types of multiple intelligence tests yielded statistically significant differences in all of the seven intelligence areas (Table 1). This supports the premise that if the students' responses to the various test items were affected by mis-perceptions of written language, then their ranking of intelligences would indicate a difference between the two types of tests, as was observed.

An alternative explanation would be that each test is measuring different attributes. The validity of each test had been a priori assumption. If both the TIMI and MIQ tests were measuring the same attributes, the observed differences could be due to the suggested differences in perceptual ability in language and pictorial interpretation, or to other possibly unanticipated factors. Since differences between the rankings of ESL students on the two types of tests has been established and tentatively attributed to differences in written language and pictorial interpretation, comparison with EFL students was needed to determine if this was a shared difference in perception, or unique to ESL students.

When the Pearson correlation coefficient was determined for paired intelligence rankings indicated by ESL students, there was a statistically significant correlation between the two types of tests in only one intelligence area — math (Table 1). It had been anticipated that there would be stronger positive correlation in all areas of intelligences. Is it only in this one area, math, that the same concepts are being measured?

The reason for the unexpectedly weak correlation between the student's rankings on the two tests needs to be explored further. Is the difference between pictorial and written ranking by ESL students due to a difference in language perception or a difference in what the tests are measuring? Are preferred intelligences similarly distributed in each language group? If they are, both tests should provide similar mean rankings for each language group if they are measuring the same thing. To investigate whether intelligences are similarly distributed, the second research question, t-tests considering ESL and EFL students as independent groups were carried out for each intelligence on each type of test.

Rankings by ESL vs. EFL Students

The second research question asked whether on a given type of test, there was a significant difference between EFL student's ranking of a given intelligence and the ranking of ESL students for the same intelligence (Table 2). It had been anticipated that the EFL and ESL results would not show a significant difference between the two groups of students. This assumption was based on similarities between the two groups in all demographic aspects except for English language ability, and the fact that individual differences would be obscured by the process of calculating the group mean.

The findings from this comparison showed that in most of the areas of intelligence the two language groups' mean ranks were similar (Relationship #2a. and 2b – Figure 2). There were only two areas where there was a significant difference between the ESL and EFL groups: in the area of linguistics, with the TIMI pictorial, and in the area of music with the written MIQ test. It is conjectured that even these differences between the language groups may not have been significantly different if the number of students in the EFL sample had been larger.

Since the ranking of the results of a given intelligence by the two language groups was generally not significantly different on a given type of test (Table 2), the preferred intelligences appear to be distributed in a similar pattern across both language groups. This comparison of language groups doesn't appear to supply the answer to why the ESL students' paired sample t-tests were not generally correlated. To address this issue and the final research question the ranking of intelligences by EFL students on each test will be discussed.

EFL Rankings on Written vs. Pictorial Tests

The third research question was, "Is there a significant difference between EFL students' rankings of their intelligences on a written test in English, and their ranking of the same intelligence on a pictorial test?" It had been anticipated that no significant differences would result when parallel written and pictorial tests were administered to EFL students. If the two types of tests were both measuring the same attributes there should be similar results between the tests in every intelligence area. This should be especially evident in the case of EFL students who would be least likely to be affected by the differences between the written test and the visual pictorial test. There was no significant difference of EFL students ranking in five of the intelligence areas on the two types of tests, but in two intelligence areas: spatial and interpersonal intelligences, the paired sample t-test showed significant differences (Table 3). This suggests that EFL students are not generally affected by the change in assessment mode from written to a pictorial form of a test. A larger sample of EFL students needs to be tested to determine whether the differences in spatial and interpersonal intelligences are typical or atypical.

Unexpectedly, correlation between the two types of tests given to just the EFL students showed no significant correlation except in the area of music intelligence. A larger sample may indicate correlations not detected for the other six areas of intelligences in this small sample. This leaves unsolved the question of whether the two tests are measuring the same attributes. If a large sample of EFL students indicated a significant positive correlation between TIMI and MIQ rankings of intelligence, this would support the validity of the two tests and also

strengthen the argument developed in this work that ESL students respond differently to written and pictorial tests.

It was thought that comparing the two types of tests would give a clearer understanding of the problem involved in ESL students' assessment. This study has found that the ESL students' ranking of each intelligence on a pictorial test is significantly different to their ranking of that intelligence on a written test (Table 1). However, there is no significant difference between the rankings by EFL students of most intelligences on each test. When compared as language groups, the EFL and ESL students' rankings of intelligences are very similar.

These findings support the proposition that ESL students respond to the written assessment used in a different way to the pictorial assessment. However clear interpretation and generalization of this observation is obscured and limited by the lack of correlation between rankings on the two instruments. Since there is a lack of correlation between the two types of tests used, further re validation of one or both instruments may be warranted.

Conclusion

The results of this study are inconclusive. It was expected that the investigation of questions regarding differences in the decoding of textual and visual materials by students with ESL backgrounds as compared to students with a native English-speaking background would give us a clearer understanding of methods for assessing ESL students. It was anticipated that ESL students would have a different understanding of questions on the pictorial test than questions on a written test. This would be evident by significantly different scores on the two types of tests. The fact that there were differences in test results between the two language groups on the two tests suggests that the original hypothesis of differences between ESL and EFL students was supported but not completely validated by this study. A future extension of this work would be to investigate by interviews whether pictorial tests more correctly represented ESL students' views.

References

Batulayan, Nerissa S. (2001). *Relationship between multiple intelligences and academic achievement of grade six pupils in Northern Luzon Mission* Silang, Cavite: Adventist International Institute of Advanced Studies.

Gardner, Howard (1983). *Frames of mind: the theory of multiple intelligences* New York: Basic Books.

Kaufman, Dorit (1996). *"Interdisciplinary collaboration in teacher education: A constructivist approach."* Stony Brook, New York, Department of Linguistics, State University of New York.

Kifer, Ken. (1994). *"On using the Nelson-Denny test for ESL students a X college."* Ken Kifer's Writing Pages. http://www.kenkifer.com/writing/ellis.html (posted April 2002).

LaCelle-Peterson, M., and Rivera, C., (1994). *"Is it really real for all kids?" A framework for equitable assessment policies for English language learners."* Harvard Educational Review 64, 1: pp.55-75. Cited by Beverly J. Irby, Sam Houston State University, Bilingual Research Journal. Spring, Summer, and Fall 1998, Vol.22, pp. 2,3, & 4.

McKeon, Denise. (1994). "*When meeting 'common' standards is uncommonly difficult."* Educational Leadership Vol 51, pp.8. http://www.ascd.org/readingroom/edlead/9405/mckeon.html (posted May 1994).

O'Malley, J.M., and Waggoner, D., (June 1984). *"Results of a U.S. survey: public school teacher preparation in the teaching of ESL."* TESOL Newsletter 18: pp.3. http://www.semlab2.sbs.sunysb.edu/Users/dkaufman/action.teacher.ed.final.html

Teele, Dr. Sue. (1994). *"Reforming the educational system to enable all students to succeed."* University of California, Riverside, Education Extension. http://www.unex.ucr.edu/education/MI/reforming.html (posted August 2000).

Teele, Dr. Sue. (1992). Teele inventory for multiple intelligences. Redlands, California, Sue Teele & Associates.

Appendix # 1

Teele (1994) summarizes her ideas regarding each of the seven multiple intelligences as follows:

Linguistic students have highly-developed auditory skills, enjoy reading and/or writing, like to play word games and have a good memory for names, dates and places. They may possess well-developed vocabularies and use language fluently and are often able to spell words accurately and easily phonetically.

Logical-mathematical students like to explore patterns and relationships and enjoy doing activities in a sequential order. They often like mathematics, experiment to test things they don't understand, enjoy opportunities to solve problems and reason logically and clearly.

Spatial students enjoy art activities, reading maps, charts and diagrams and thinking in images and pictures. They are able to visualize clear images when thinking about things, and can complete jigsaw puzzles easily.

Musical students are sensitive to the sounds in their environment, enjoy music and prefer listening to music when studying or reading. They appreciate pitch, rhythm and timbre and often sing songs to themselves.

Bodily-kinesthetic students process knowledge through bodily sensations and use their bodies in differentiated and skilled ways. They need opportunities to move and act things out, and tend to respond best in classrooms that provides physical activities and hands-on-learning experiences.

Intrapersonal students prefer their own inner world. They like to be alone and are aware of their own strengths, weaknesses and inner feelings. They often have a deep sense of self-confidence, independence and a strong will, and motivate themselves to do well on independent-study projects. They may respond with strong opinions when controversial topics are being discussed.

Interpersonal students enjoy being around people, have many friends, and social activities, and learn best by relating and participating in cooperative group environments. These students express empathy for the feelings of others and respond to their moods and temperaments.

(adopted from Teele 1994)

Jean Streifling
is a PhD student in the Department of Educational Studies
Adventist International Institute of Advanced Studies