Interactive and Affective Behaviors of Teaching Assistants in a First Year Physics Laboratory

by

Zahra Hazari

A.W. Key

and

John Pitre

University of Toronto

Introduction

Training teaching assistants (TAs) to teach the science and technology that are integral to the undergraduate laboratory is difficult; to instill the affective and human dimensions of teaching is even more so. To inform and support this task in our TA training program, we investigate some behavioral characteristics that lead to student satisfaction. Our inspiration was the seminal work of Murray (1977, 1980), who studied the correlation of classroom behaviors of social science lecturers with student ratings. He found that the frequency of behaviors such as 'speaks expressively', 'tells jokes or anecdotes', 'enthusiastic', and 'shows strong interest in subject matter' were significantly different for high, medium, and low rated groups of lecturers

The frequency and nature of the interaction between teacher and student, as well as the affective behaviors of the teacher are crucial elements in the instruction process. McLeod (1996), for example, found that an increase in interpersonal exchanges often facilitates more effective periods of learning, while Pankratz (1967) demonstrated a strong correlation between the student rating of teachers and the length of the teacher-

student interaction. In addition, Frymier (1993) found that students beginning the term with either low or moderate motivational states were found to have increased levels of motivation later in the term when exposed to a highly immediate teacher. Henderson et al. (2000) found strong correlations between students' perception of interpersonal teacher behavior and student attitude in biology laboratories.

In terms of affective behaviors, Haskins (2000) found that teachers use behaviors such as vocal variation (pace, inflection, volume, vocal expressiveness, etc.) and visual variation (facial expressions, smiling, eye contact, gestures, movement, etc.) to help communicate subject matter. McCrosky and Richmond (1992) confirm that these verbal and non-verbal behaviors enhance student learning and students' liking for the subject matter. She and Fisher (2002) found that students' attitude and cognitive achievement in science were higher when students perceived their teacher as giving more encouragement and praise, more nonverbal support, and being more understanding and friendly. Finally, McDowell (1993) found that even 60-80% of TAs in his study rated 'friendliness', 'communicator image', 'impression leaving', 'attentiveness', and 'animated' more positively than other style variables.

Arreola (1995) has examined many of the criticisms leveled against student evaluations of teachers. Student evaluations have been criticized on the grounds that students are not sufficiently mature or knowledgeable to make reliable judgements. This criticism has been effectively refuted; the studies of, for example, Costin et al. (1971), Gillmore (1973), and Hogan (1973) find consistently high correlations between student ratings of the same instructors and courses from year to year. Another criticism is that student ratings of teachers are little more than a popularity contest in which students

reward teachers who exhibit immediacy and warmth. However many studies by Aleamoni (1976), Frey (1978), and Arreola (1983), among others, show that students can well distinguish between instructional effectiveness and affective behaviors that lead to high popularity. Some observers have complained that student ratings of teachers are little more than a popularity contest in which students will reward teachers who give them high marks. However, in his survey of the extensive research on this question Arreola (1995) concludes that 'the belief that student ratings are highly correlated with their grades is not supported by the literature'.

We designed two instruments to investigate some of these questions; the first recorded the frequency and the nature of the interactions of our TAs with their students, and the second provided student evaluations of the TAs. We were interested in determining whether students preferred to initiate the interactions with their TA, or to wait for the TA to do so. We also wanted to determine the affective behaviors that were highly rated by students. Our student evaluation questionnaire followed closely the work of Murray (1977). It included a set of questions that measured the students' evaluation of the quality of the *Assistance* they received, the *Fairness* of their TA, and the *Influence* the TA was perceived to exert in instilling positive attitudes towards physics. Since most of the students in our study will study no more physics after graduating from the first year course, we were particularly interested in the TAs' ability to influence their attitudes to physics that will presumably last into later life.

The Study

In our largest first year physics course, as in many introductory courses, the lecture section is complemented by a laboratory session in which the students meet in

small groups under the supervision of a graduate student employed as a laboratory TA (TA). The small class size, the closeness in ages of the students and their TA, and the easy accessibility of the TA leads to the expectation that these sessions will have an important influence on students' appreciation of the laboratory and of experimental physics.

Student motivation for physics in this course is not high since most of the students take physics only as an entry to life science programs. We hypothesize that TAs who have more frequent and more empathetic interactions with their students will promote the highest satisfaction with the course and the most positive attitudes towards physics. We selected nine highly qualified TAs for our study; all were senior doctoral students in good standing who had taught the course at least once before.

Our data was collected using two instruments; the Student Evaluation of the TA (SETA) and a TAs' Behavior Catalogue (TABC) (see appendix). The 15-item SETA provided student ratings in three main categories: the *Assistance* provided by the TA, the *Influence* of the TA on student attitudes to physics, and the *Fairness* of the TA. The *Total* student evaluation that included all three categories was used to classify the TAs into **High**, **Moderate**, and **Low** levels of student satisfaction.

The second instrument, the TABC, was an observational instrument we designed to measure two types of interactive behavior of TAs. The first type was 'interactive behavior' measured by the frequency of interaction and identified as either Student-Initiated (SI) or TA-Initiated (TI). The second type was 'affective behavior' (smiles, eye contact, etc.) measured using a 7-point Likert Type Scale. A later factor analysis of the affective behaviors measured by the TABC yielded three categories; 'Amiableness',

'Calmness/Clarity', and 'Aggressiveness'. Six observers completed the TABC for each of the nine TAs. Figure 1 summarizes the details of the study.

Finally, we administered a short multiple choice Pretest and used the results of the multiple choice laboratory test taken by all students at the end of the course to check that the students in our study did not differ significantly in preparedness or final achievement.

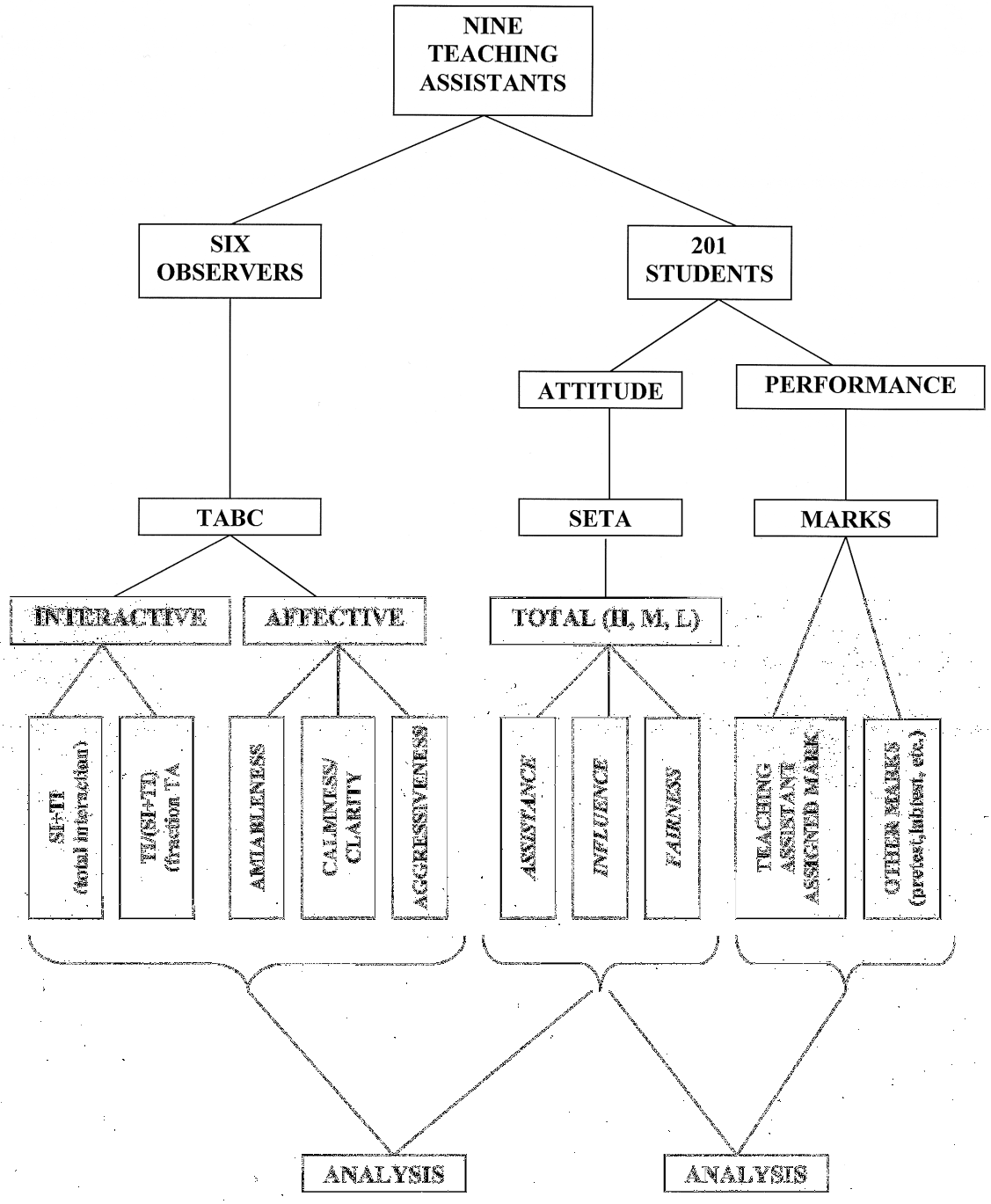

Figure 1. Schematic flow chart of the Study.

The Laboratory

Approximately 1500 students from four different physics courses attended the first year physics laboratory in the fall of 1999. Students attended three-hour lab sessions

every other week for a full academic year. 44 TAs supervised the students in groups of 10 to 18, as they guided them through a common set of experiments. The aim of the laboratory was to teach experimental science; to this end the TAs worked closely during the session with the students in their group, paying attention to experimental technique and the acquisition and analysis of data. The TAs marked the students' work as recorded in their laboratory notebooks and evaluated their performance in the laboratory. Most TAs supervised two groups of students.

The laboratory was quasi-traditional in nature. Although many of the experiments were verification experiments (the value of the acceleration of gravity, Ohm's Law, etc.), the laboratory manual did not present the experiments in step-by-step form. It merely presented the background theory, indicated the various different aspects the students were required to study through experiment, gave hints, and asked leading questions. The students were required to do a good deal of thinking for themselves in order to conduct the experiment and provide a detailed laboratory report. Students were also required to keep track of all sources of error and account for them with stringent error analysis. Six hours were allotted to each experiment.

Method

Pretest & Labtest

We administered a pretest to all students in the laboratory at the start of the academic year (September 1999). The pretest had six multiple-choice questions on basic laboratory skills such as interpreting or extrapolating information from graphs. At the end of the academic year (April 2000) students sat a written multiple-choice laboratory test that probed their understanding of the experiments they had performed. We used this

as a measure of student learning in the laboratory in the absence of a practical test that might have had more direct relevance to the teaching of the TAs.

Systematic Observation

Pilot testing for this study was conducted in the summer 1999 session and the beginning of the fall 1999 session of the first year physics laboratory. Our measurement instrument (TABC) was modified several times leading to a final version that was easy to use, measured specific interactive and affective behaviors of interest, and provided good inter-rater reliability (see Table 1, below).

During the month of November 1999, systematic observation was conducted using the TABC. In order to reduce sources of variation due to beginner anxiety or ignorance of the experiments we chose qualified TAs who had previously taught this laboratory. Our resources allowed us to study nine TAs out of the 16 who fit the criteria. The final choice was made on the basis of compatibility of schedules between the hours of the laboratory sessions and the observers. As might be expected from a total population containing only nine females out of 44, TAs chosen by this method were exclusively male. We observed each TA for approximately two hours of three different laboratory sessions (approximately six hours of total observation time per TA) with at least two different groups of students. Each TA was observed by all six observers, working in a different pair, for each of the TA's three sessions.

Student Evaluations

To determine the level of student approval and satisfaction we developed an evaluation instrument, SETA. As is shown in Appendix 1, questions concerning the students' attitudes to Physics and to the Laboratory were also included. A total of 657

evaluations were collected, 201 of which were from students supervised by one of the nine subject TAs.

Student Interviews

In order to obtain qualitative as well as quantitative data and to modify and validate the SETA, we conducted several student interviews (March 2000). The tape-recorded interviews were conducted with groups of two students (whenever possible with one male student and one female) chosen at random from one of the groups of each of the nine subject TAs. The interviews focused on the students' reflections and responses to the SETA items as well as allowing them to respond to more open-ended questions regarding their TA's behavior. The interviews thus provided us with qualitative data for each of the nine TAs we were studying (see Appendix 2).

The Measurement Instruments

A TAs' Behavior Catalogue (TABC)

There exist a wide variety of instruments that have been designed to observe teaching. The early work of Pancratz (1967), for example, concentrated on the content and frequency of classroom interactions between teacher and student. More recently, the Reformed Teaching Observation Protocol (RTOP) allows mathematics and science teachers to reform their teaching by having their lessons observed and scored on a variety of measures from lesson design to classroom culture using RTOP (MacIsaac and Falconer, 2002). Similarly, the Horizon Research Group's protocol rates mathematics and science lessons on lesson design, implementation, mathematics/science content, and classroom culture through structured observation and provides an overall rating of the lesson (Weiss et al., 2003). The goal of RTOP and the Horizon's protocol is to rate

lessons on various measures thereby determing the quality of instruction and identifying the areas needing reform.

Other instruments have also been designed to measure teacher behavior or classroom environment in the form of questionnaires rather than observation protocols. For example, the Questionnaire on Teacher Interaction (QTI) measures students' perceptions of interpersonal teacher behaviour and the Science Laboratory Environment Inventory (SLEI) measures students' perception of the laboratory environment (Henderson et al., 2000). The Teacher Communication Behavior Questionnaire (TCBQ) has been used to evaluate both students' and teachers' perceptions of science teachers' interpersonal communication behaviors (She and Fisher, 2002). Although the measures used on QTI and TCBQ are similar to our TABC measures, the TABC is an observation instrument and not a questionnaire. But unlike RTOP and the Horizon's Protocol, our purpose for observation is not to rate the instruction but merely to catalogue the behaviors exhibited by the TAs.

The questions on RTOP and similar protocols require a teacher whose range of movement and behavior is fairly well-defined within a limited physical space, whose lesson plan is pre-prepared and structured, and whose speech can be clearly heard. The atmosphere in our large laboratory is quite inimical to such subtle observation. Also, these protocols and others we investigated address many issues that are not relevant to the instruction of a TA in a laboratory setting. Indeed trial runs with questions similar to those of these protocols yielded unacceptably low reliability coefficients. Given these difficulties, we designed the TABC for the explicit purpose of systematically recording

certain interactive behaviors of the TAs that could be clearly and unambiguously observed in this complex environment (see Appendix 3).

The first section of the TABC records a simple count of the interactions between a TA and one or more students. Each interaction is classified as either a Student-Initiated Interaction (SI) or a TA-Initiated Interaction (TI). SI is identified by a verbal initiation or a gestured prompt (for example, raising a hand) on the part of a student; similarly for TI on the part of the TA. A third category recorded on the TABC is No Interaction (NI) which classifies time periods of 20 seconds or more where the TA is not interacting with students[i]. The two main variables used in analyses of data from this section were (SI+TI), which measures the total number of TA interactions per student per ten-minute period, averaged over the student groups of each TA, and TI/(SI+TI), which measures the fraction of these interactions that were initiated by the TA.

The second section of the TABC measures 24 different inter-personal or affective behaviors of the TA. The behaviors include categories such as enthusiasm, clarity, eye contact, smiles, frowns, etc. Observers watched for these behaviors throughout the 2-hour observation period and at the end of the period they gave a score for each of the behaviors on a 7-point Likert Type Scale. The three main variables used in analyses of data from this section were 'Amiableness', 'Calmness/Clarity, and 'Aggressiveness', constructed from the individual affect categories. The description of the three variables and how we arrived at them is described below in the section entitled G Study.

Reliability of the TABC

Inter-rater Reliability

The TABC was considerably modified and simplified during a pilot study to ensure inter-rater reliability – i.e. that different observers, observing the same TA on the same occasion would agree by more than chance agreement. The results for Cohen's kappa (Cohen, 1960) for each TA are shown in table 1; values of 0.70 are considered satisfactory.

Table 1.
Inter-rater reliability for pairs of observers.

TEACHING ASSISTANT	COHEN'S KAPPA
1	0.80 0.71 0.84
2	0.74 0.83 -
3	0.88 0.87 0.83
4	0.84 0.82 0.83
5	0.86 0.78 0.81
6	0.95 0.78 0.87
7	0.64 0.64 0.66
8	0.69 0.78 0.72
9	0.74 0.87 0.77

G Study

A Generalizability Theory study (Shavelson and Webb, 1991) is a measure of reliability that indicates how well the results of the study can be generalized (are consistent), despite multiple sources of variability. In our case these sources of variability are different occasions of measurement, different observers, and extraneous unknown effects (such as a change in environmental conditions from one occasion to the next, a momentary lapse of the observers' attention, etc.). The magnitudes of variability

due to each source are used to compute a G coefficient, which measures how well the results can be generalized across all sources. For example, how well performance on one occasion can be generalized to performance on all occasions, or how well observations by one observer can be generalized to all observers. A G reliability coefficient value of 0.7 or greater is taken to be acceptable (Nunnaly, 1978). Table 2 lists the percentage variances for the different sources of variability in our experiment[ii] for the five TABC measures: the first two, (SI+TI), TI/(SI+TI)), from the interaction behaviors section and the last three ('Amiableness', 'Calmness/Clarity', 'Aggressiveness') from the affective behaviors section. The G coefficients for all five measures are greater than 0.75, indicating that all of the five measures are generalizable over occasions, observers, and extraneous effects.

Table 2.
G Study Results.

SOURCE	SI+TI	TI/(SI+TI)	Amiableness	Calmness/ Clarity	Aggressiveness
TA	58%	53%	81%	27%	61%
Occasion	31%	32%	5%	17%	8%
Observer	0%	4%	11%	48%	27%
Extraneous	10%	11%	2%	8%	4%
G coefficient	0.85	0.83	0.97	0.76	0.94

The variance due to TAs dominates for four of the five measures with 58%, 53%, 81%, and 61% of the total variance in these cases being due to differences in the TAs. The one case where the variance due to TA does not dominate is the measure Clarity/Calmness (27%) which indicates that this measure is not as reliable for distinguishing between TAs as the other four.

For (SI+TI) (total number of interactions) and TI/(SI+TI) (ratio of TA initiated interactions to total interactions) the variance due to occasion (31% and 32% respectively) is much greater than the variance due to observer (0% and 4% respectively). This result is expected since both SI and TI are frequency counts, which should not differ significantly for reliable observers but may very well differ on different occasions depending on the students in the class, the stage of the experiment reached, etc. For 'Amiableness', 'Calmness/Clarity', and 'Aggressiveness', the variance due to observer (11%, 48%, and 27% respectively) is much greater than that due to occasion (5%, 17% and 8% respectively); indeed the affective behaviors that describe a given TA might be expected to vary less from occasion to occasion than the subjective judgement of these behaviors by different observers.

Validity of the TABC

The validity of an instrument depends on whether it measures what it is intended to measure. The question of validation of the TABC arises in the identification of an interaction and the differentiation of interaction types by the observers. To ensure that our observers were actually recording interactions and categories that fit our definition, we conducted training periods using verbal discussion and group observation to clarify our definitions by pointing out explicit examples of interactions and interaction types. In this way we confirmed that all observers were recording the same measure of an interaction. The affective behaviors scale of the TABC was similarly validated through extensive clarifications between the six observers on all the 24 affect categories as well as on the seven points of the Likert Type Scale for each category.

Student Evaluation of the TA (SETA)

The SETA consisted of 15 items on a 5-point Likert Type Scale (see Appendix 1). The first 14 items focus on questions about the TA's knowledge and performance, inter-personal attributes, and influence on student attitudes. The last item, question 15, asks students for a rating of the Laboratory independently of their rating of their TA.

Reliability of SETA

An analysis of internal consistency was used to establish the reliability of the SETA. Internal consistency reliability indicates the extent to which items correlate among themselves (Sax, 1997). The reliability coefficient we used was coefficient alpha (Sax, 1997); a value of alpha greater than 0.7 establishes reliability (Nunnaly, 1978). The SETA scale turned out to be highly internally consistent with a coefficient alpha of 0.95.

Validity of SETA

The validity of the SETA is based on student interviews that focused on students' reflections and responses to the SETA items. Through these interviews, the SETA items were modified and reworked to remove ambiguities in wording and interpretation and to ensure that each item measured what we intended it to measure. In addition, a factor analysis was also performed on the SETA scale results. A factor analysis identifies the minimum number of factors that account for test variance (Sax, 1997). In our case, the factor analysis resulted in a one-factor solution, indicating that each item on the SETA scale was highly correlated. This corresponded to our intention that each item, which measured some aspect of the TA's behavior, indeed provided a consistent contribution to the overall evaluation of the TA.

Results & Discussion

The SPSS statistical package and guide (SPSS Guide, 1999) were used to provide the analyses discussed in this section.

Pretest and Labtest

An Analysis of Variance (ANOVA) of the pretest scores alone gave non-significant results across the subject TAs. This indicates that all the subject TAs' groups were at a similar level of student ability at the start of the study, thus providing a good starting base. An ANOVA of the laboratory test and of student test gain (from pretest to laboratory test) indicated no significant differences across the nine subject TAs. Hence, the students in our subject TAs' groups were also at the same ending level in terms of learning. In addition, there was no correlation between the student evaluations and student test gain. Thus, it appears that student ratings were not affected by student learning as measured by these tests.

It should be noted that we have reservations about the level to which these tests measure deeper student learning in the laboratory. In undergraduate laboratories that mimic real-life experimental research, there are many complex demands made on students and many types of learning that are not well quantified using standard tests; the ability to understand unfamiliar equipment, to gracefully negotiate complex instructions, to design elegant and efficient experimental solutions to unforeseen and unforeseeable practical problems, etc. The teaching of experimental physics is a much more complex activity than lecturing or tutoring - at least in the complex, many faceted experience of which our laboratories are good examples - and the identification of meaningful laboratory-specific learning in such large classes is very difficult, if not impossible. The

pretest and post-test (laboratory test) available in our study were certainly inadequate for such a purpose. In addition, the first year physics laboratory occupies a very small part of the typical course load of five full year courses, at least two of which have their own laboratory (e.g. Chemistry and Biology); the physics laboratory counts for only 20% or so of the overall mark in the physics course. It would be surprising if the physics laboratory had a clearly measurable impact on students' overall learning. For this reason we chose to concentrate on student satisfaction.

Student Evaluation of the TA

A factor analysis of the SETA data resulted in a one-factor solution, indicating the consistency of all the items. We therefore define the *Total* evaluation to be the sum of the scores on all items of the SETA except the last, which evaluated the laboratory alone; this *Total* was used as a convenient single measure of the student evaluation of the TA. However, the questions on the SETA fall naturally into categories that help clarify the later analysis: when we force multi-factor solutions, it becomes evident that two of the questions are inconsistent with the rest. Omitting these two questions, we obtain a solution with three major factors shown in Table 3. We label these three factors *Assistance* (in guiding the student), *Influence* (on the student's attitude to physics), and *Fairness* (in marking and treatment of the student).

Table 3.
Category groupings for the SETA.

FACTOR	CATEGORY GROUPINGS
Assistance	Availability, Usefulness, Communication Skills, Knowledge, Overall.
Influence	Enthusiasm, Influence, Stimulation, Inspiration.
Fairness	Fairness in Marking, Friendliness, No Favourites.

An ANOVA done on the *Total* student evaluation scores resulted in a significant difference ($F(8,190)=3.732$, $p<0.001$) across the nine subject TAs. The Tukey Post Hoc test resulted in the separation of the subject TAs into three homogeneous subsets with three TAs per subset. These three subsets were labeled **High**, **Moderate**, and **Low**, to distinguish the levels of student satisfaction with the TA.

SETA question 15 asks: 'Independently of my rating of my TA, (what is) my overall rating of the First Year physics laboratory….' The correlation of the *Total* evaluation with the answers to this question was positive and highly significant for both the nine subject TAs ($r=0.320$, $p<0.001$, $N=201$) and the entire population of 44 TAs ($r=0.367$, $p<0.001$, $N=614$). Thus students' appreciation of the laboratory and experimental physics is directly related to how they perceive their TA.

There is some evidence that students tend to evaluate more highly those teachers from whom they receive higher grades (Lewis, 1998). Indeed we observe such an effect; however, it is not a simple one. While a correlation of the *Total* evaluations of the TAs with the mark that they assigned was positive and significant for all 44 TAs, it was not significant for the nine we studied[iii]. A clearer picture emerged when the three categories of the SETA were examined separately. The correlation of the TA assigned mark with the *Fairness* category was indeed significant and positive for both groups of TAs[iv]. However, correlations with *Assistance* and *Influence* were not significant for either group. Apparently students can discriminate between their TA's effectiveness in providing assistance in the laboratory and in positively influencing their attitudes to physics from the mark that their TA assigns them. This result is in good agreement with the results reported in Arreola (1995)

Interactive Behaviors measured by the TABC

ANOVAs

The **High**, **Moderate**, and **Low** rank subsets of TAs, distinguished by student evaluations, were compared in terms of the interactive behaviors they exhibited[v]. We observed significant differences ($F(2,51)=17.313$, $p<0.001$) between the three groups in terms of the fraction of total interactions that were TA-initiated (TI/(SI+TI)). Also, the total number of interactions (SI+TI) was significantly different ($F(2,51)= 4.479$, $p<0.01$) between the **Low** group and one that combined both **Moderate** and **High** groups[vi]. There was no significant difference, however, between the **Moderate** and **High** group. The bar charts in Figure 2 illustrate these results.

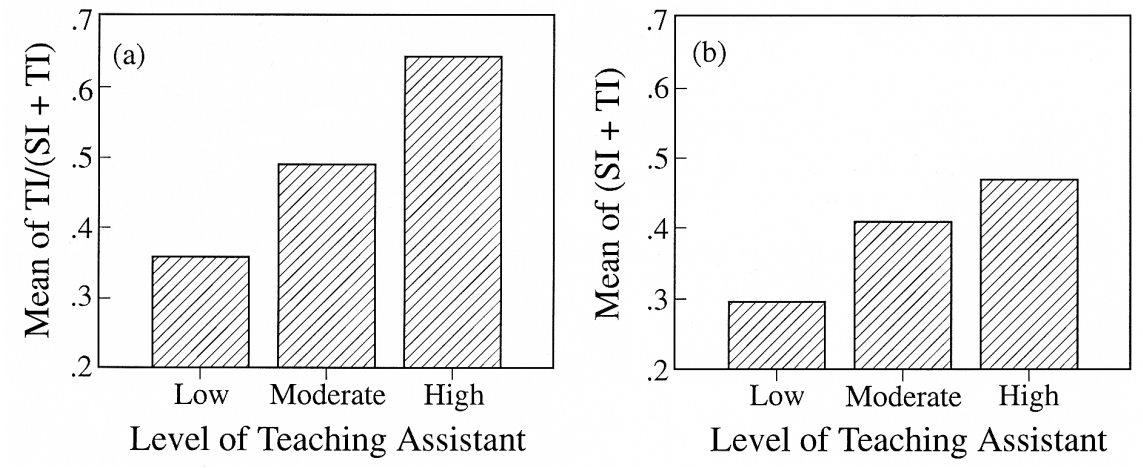

Figure 2. Bar charts of mean values of (a) TI/(SI+TI) and (b) SI+TI for **High**, **Moderate**, and **Low** rank demonstrators.

Correlations

Correlations were measured between the interaction variables (SI+TI, TI/(SI+TI)) and the various categories of the SETA (*Total*, *Assistance*, *Influence* and *Fairness*). Table 4 shows the correlation coefficients.

Table 4.
Pearson correlation coefficient (r) between interaction and SETA categories.

CORRELATIONS	*Total*	*Assistance*	*Influence*	*Fairness*
N	9	9	9	9
TI/(SI+TI)	0.799**	0.675*	0.875**	0.791*
SI+TI	0.596*	0.613*	0.618*	0.345

*significance level $p<0.05$, ** significance level $p<0.01$

Figure 3 shows the regression plots of the interaction variables with the *Total* category. The correlation coefficients between the *Total* evaluation and the interaction categories are positive and significant. As suggested by previous results, the fraction of TA-initiated interactions is more strongly correlated with the SETA categories than the total number of interactions; this is most evident for the *Influence* category. We might guess that this latter variable is related to the attention, and, by implication, the concern that the TA shows towards the students. This concern might be expected to be important in influencing the students' attitude towards physics.

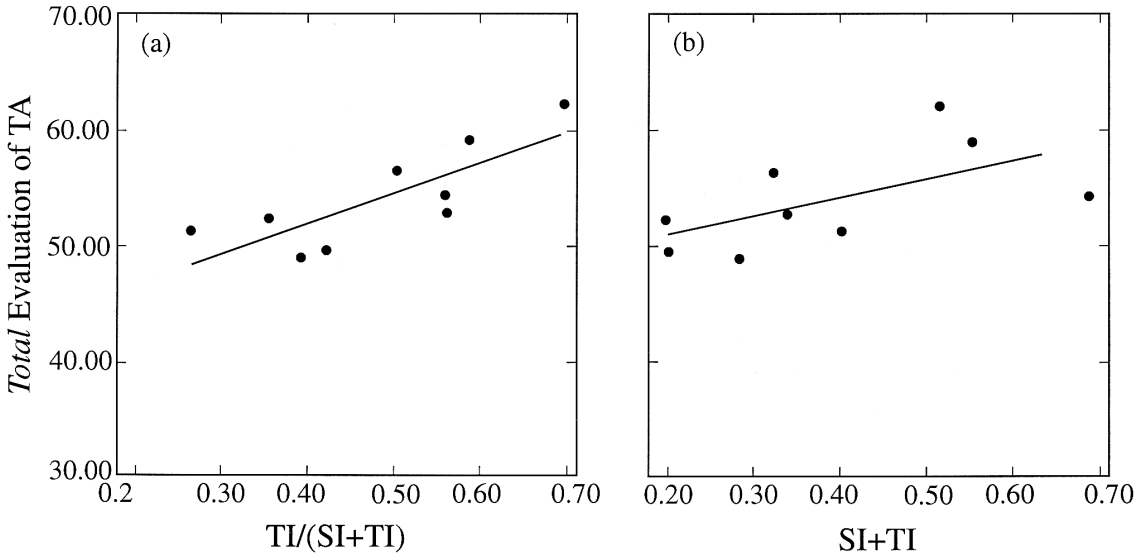

Figure 3. Regression plots of average *Total* student evaluation of demonstrator versus (a) TI/(SI+TI) and (b) SI+TI.

Affective Behaviors measured by the TABC

Factor Analysis

A factor analysis of the 24 categories measured in the affect section of the TABC resulted in a six-factor solution, with three of the factors containing only a single category; two of these did not correlate well with any of the other factors. Since the scree plot[vii] indicated a three-factor solution, this solution was forced and the two latter categories were removed. The final category groupings, which we name 'Amiableness', 'Calmness/Clarity, and 'Aggressiveness', are shown in table 5.

Table 5.
Category groupings for the affective measures of the TABC.

FACTOR	CATEGORY GROUPINGS
'Amiableness'	Agreeableness, Concern, Cheerfulness, Energy, Enthusiasm, Warmth, Excitement, Jokes/Chats, Praises, Smiles, Eye Contact, Uses Names, Pitch, Pace.
'Calmness/Clarity'	Calmness, Clarity.
'Aggressiveness'	Criticizes, Frowns, Physical Contact, Shows Frustration, Volume.

ANOVAs

The **High**, **Moderate**, and **Low** rank TAs were compared in terms of these three factors. We found that 'Amiableness' was significantly different ($F(2,51)=18.225$, $p<0.001$) between the **High** and a combined **Moderate** and **Low** groups, but not between the **Moderate** and **Low** group. We found a similar result for 'Calmness/Clarity' ($F(2,51)=3.436$, $p<0.05$). Thus, the highly evaluated group of TAs were significantly more 'Amiable' and 'Calm and Clear' than the less highly evaluated groups of TAs. The bar charts in Figure 4 illustrate these results. The data on 'Aggressiveness' yielded no statistically reliable result.

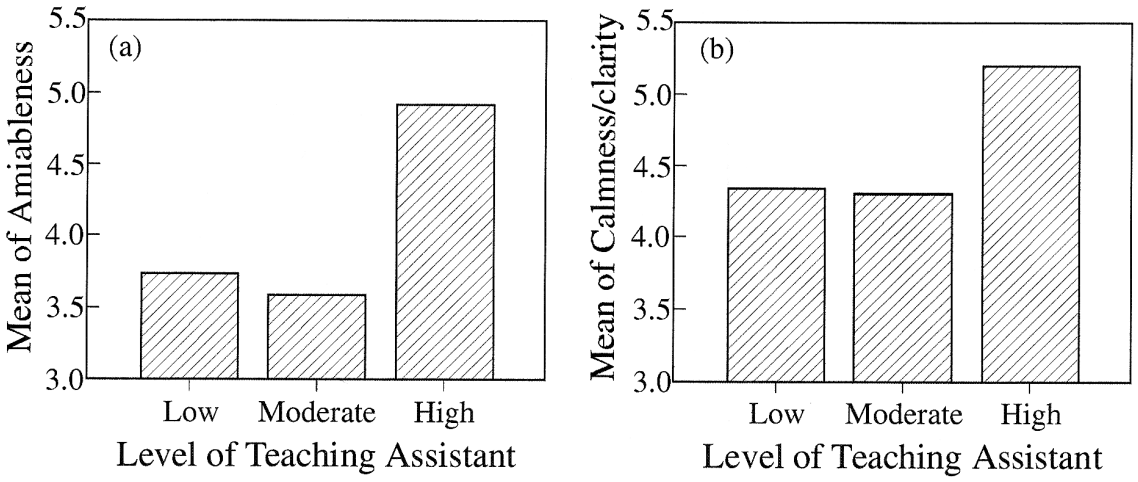

Figure 4. Bar charts of mean values of (a) Amiableness and (b) Calmness/Clarity for **High**, **Moderate**, and **Low** rank demonstrators.

Correlations

Correlations of the SETA categories with 'Amiableness' and with 'Calmness/Clarity' are shown in table 6.

Table 6.
Pearson correlation coefficient (r) between affective and SETA categories.

CORRELATIONS	*Total*	*Assistance*	*Influence*	*Fairness*
N	9	9	9	9
'Amiableness'	0.751*	0.703*	0.801**	0.658*
'Calmness/Clarity'	0.522	0.642*	0.290	0.204

*significance level $p<0.05$, ** significance level $p<0.01$

Figure 5 shows the two regression plots. The 'Aggressiveness' variable showed no significant correlation with any of the SETA categories. The range of behaviors gathered under the title of 'Amiableness' are clearly a valuable asset to a TA in achieving high student ratings, and in influencing their attitudes towards physics. 'Calmness/Clarity' are also valuable attributes, but of less importance. These results strongly indicate that the TA's affective demeanor plays a vital role in student satisfaction and enjoyment in the laboratory.

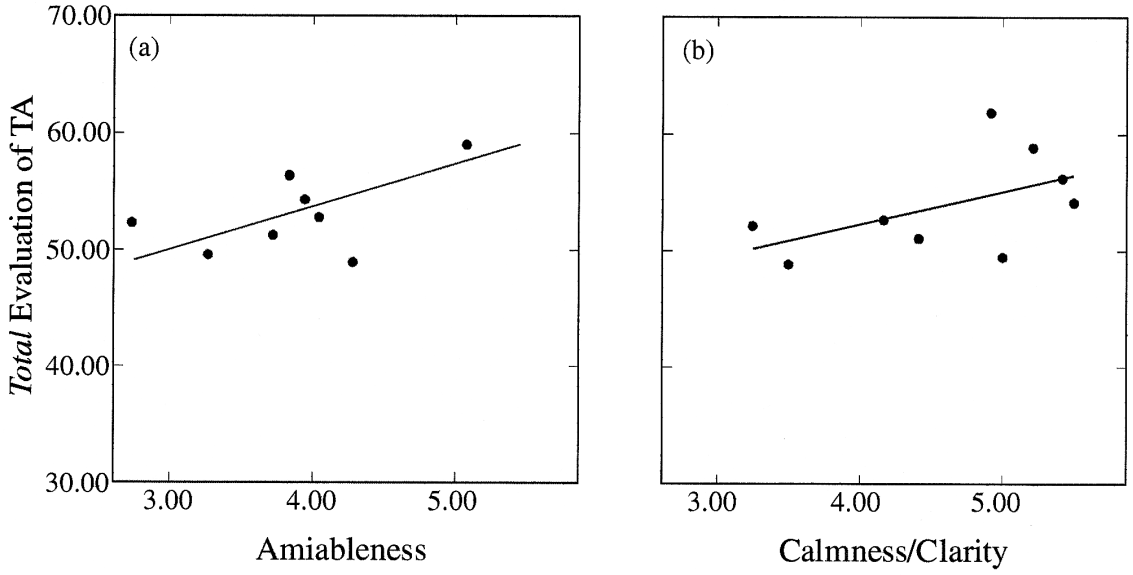

Figure 5. Regression plots of average *Total* student evaluation of demonstrator versus (a) Amiableness and (b) Calmness/Clarity.

Student Interviews

Selected excerpts from the interviews of students whose TAs are in the **High**, **Moderate**, and **Low** ranks are quoted in Appendix 2. The excerpts include all the strongly positive and negative statements made by students pertaining to our measured variables (SI, TI, and the categories of table 5). As can be observed, the students' comments make similar distinctions between the three groups of TAs to those that we identified through statistical analyses. We discovered that students appreciated TAs who exhibited behaviors that corresponded to our affective categories of 'Amiableness' and 'Calmness/Clarity'. They also liked TAs who initiated interactions, since students believed that their own questions were not sufficiently important to disturb their TA whom they perceived as being very busy. This point was elaborated through the student interviews.

Summary of Results

We have designed a measurement instrument (TABC) to measure some of the interactive and affective behaviors of laboratory TAs in a large first year physics course. The instrument is simple to use, produces good inter-rater reliability, and discriminates well between TAs exhibiting different behavior patterns. Using the TABC we measured two sets of observationally reliable measures. The first counted interactions between TAs and students, distinguishing between student-initiated and TA-initiated interactions. The second recorded observers' judgements about a variety of affective attributes that were exhibited during these interactions.

We also developed and administered a 15-item student evaluation questionnaire (SETA) that asked students to evaluate a variety of different attributes of their teachers.

An ANOVA of the student evaluations grouped the TAs into **High**, **Moderate, and Low** rank. As expected, we observed no correlation between students' evaluation and changes on a pretest and the labtest.

We found that the SETA provided a measure of TA effectiveness in terms of the TA's ability to be helpful and to provide a positive influence on their students' attitudes to physics. In order to be useful, student evaluations should be independent of the grade the student receives from the TA. When we apply factor analysis and logical grouping to the questions in the SETA, we observe that the student responses fall into three clear categories. It is only in the category that asks students about their TA's fairness that the influence of their received grade is important. However, the student responses to questions concerning other attributes of their TAs that we evaluate highly - the ability to positively influence students' view of experimental physics, and their usefulness as guides - showed no correlation with the grade assigned by the TA. In addition, we observed a strong correlation between the evaluation of the TA and the students' appreciation of the laboratory and of experimental physics. So, in the absence of more objective tests, it is in the areas of *Influence* and *Assistance* that we find a measure of TA effectiveness.

ANOVAs of the interactive and affective measures obtained from the TABC between the **High**, **Moderate**, and **Low** ranks, and correlations between these measures and the student evaluations revealed some interesting patterns. Highly evaluated TAs initiated more of the interactions with students, and had a higher frequency of total interaction than did the TAs who received lower evaluations. Further, a range of affective dimensions that we defined as 'Amiableness' was highly valued by the students,

as was that of 'Calmness/Clarity'. The students in our study responded positively to TAs who were proactive in initiating interactions. TAs who exhibited interactive immediacy helped diminish student apprehension and were considered by the students to be more effective teachers.

Discussion

The possible influence of students' grades on student evaluations has generated considerable controversy; a good discussion can be found in Arreola (1995). In our evaluations, evidence of such influence appeared only for those questions that were related to the perceived fairness of the TA. We investigate this point in more detail in another paper. Our conclusion is that, given carefully designed evaluation questions, students are quite capable of evaluating their teachers independently of the grade they receive.

We speculate that undergraduates in very large first year classes are hungry for assistance from, and personal interaction with their teachers, yet often feel too timid to approach them. A TA who initiates interactions in the laboratory relieves the student of this responsibility, and implicitly indicates their concern for the students' welfare and progress. The fact that students appreciate such TAs is hardly surprising. What is interesting is that the frequency of these TA-initiated interactions increases the influence that TAs have on their students.

Our results show that both the level of proactive exchange by a TA with students in the undergraduate laboratory and also the TA's affective demeanor positively influence the students' attitudes toward science as presented in the laboratory. This, in turn, leads to a more positive attitude toward the laboratory and to physics in general,

findings that are consistent with other studies (McCrosky and Richmond,1992; Henderson et al., 2000; She and Fisher, 2003).

The importance of the affective domain is often overlooked by physics educators who focus primarily on the cognitive domain. For students in the life sciences who typically take only one year of introductory physics, the level to which they will understand physics is superficial. However, the attitude and interest they take out of the course will impact them far more and may motivate them to further study of physics, be it formally or independently. Alsop and Watts (2003) write, “There is far more to science education than cognition...feelings of enthusiasm, confidence and zeal are equally powerful motivators, so that learners are swept up in a flow of eagerness to learn...Critics of the affective domain will often claim that such considerations focus on making students feel good at the expense of being educated. On the contrary, we argue that affect surrounds cognition." In fact, Laukenmann et al.'s (2003) results show that well-being and interest, as cognitive emotional constructs in physics instruction, play a significant role in achievement.

The implications of this study for the training of TAs - and other teachers of first year classes - are clear. Given sufficient knowledge of the material, TAs must be encouraged to be proactive in interacting with their students, and to pay attention to affective issues; friendliness, appropriate use of encouragement and language, exhibitions of interest or enthusiasm. We are in the process of developing such training exercises in the department.

Acknowledgements

We are very grateful for the strong support, both moral and financial, of the Department of Physics. We also wish to thank Dr Barry McQuarrie, Ms Michelle Ladd, and Dr Nagina Parmar for their careful observations.

Appendix 1. The Student Evaluation of TA form (SETA)

All students in the laboratory were asked to rate each of the following 15 statements on a 5-point scale from Very Poor to Very Good, and to record their responses on a mark-sense computer card.

The promptness of my TA in marking and returning my lab notebook is:
The fairness of my TA=s marking of my lab notebook is:
The availability of my TA to provide assistance during the lab session is:
The availability of my TA to provide assistance if I needed it outside the lab session is:
The usefulness of the assistance that I receive from my TA during the sessions is:
The communication skills (i.e.comprehensibility) of my TA is:
The friendliness and approachability shown by my TA to me personally is:
The fairness of my TA as a teacher who has no favourites in the group and treats all students equally is:
The energy and enthusiasm of my TA in the lab is:
Independently of the structure of the lab, the (positive) influence that my TA has had on my attitudes to physics has been:
My TA=s knowledge and understanding of experimental physics as displayed in the lab is:
The ability of my TA to stimulate me to think about my lab work is:
Independently of my mark, the ability of my TA to inspire me to do my best work in the lab is:
My overall rating of my TA as a lab teacher is:
Independently of my rating of my TA, my overall rating of the First Year Physics laboratory is:

Appendix 2. Student Interviews

HIGH
'He keeps tracking our process'
'…its helpful (the initiation of the interaction by the TA)…its really good'
'Whenever you call him, he comes within a minute'
'If we don't approach him, he'll be sure to come up to us…its good…I find it hard to ask (for help)'
'We couldn't get through without his help'
'He tries to always make himself available'
'He praises' and 'He is not critical…he teaches in a positive way'
'He criticizes but in a joking kind of way; you never feel bad afterwards'
'He is very friendly; he stays after and talks to us'
'The enthusiasm he has is really nice…sometimes (in the lab) its really boring, complicated things; the way he explains with enthusiasm makes it nice'
'He is really an inspiring person through his enthusiasm'
'He'll make sure you understand'
'He has improved my attitude'
'I think he cares'
'…I've been stimulated to enjoy it (physics)'
'He does the job really good'

MODERATE
'He comes by to see how we're doing…he doesn't do it too much though'
'The lab is hard and I find I need more help and if he comes by more frequently then I think its better off for us'
'He is sometimes busy and I go ask 'when your done please come by' and that happens a lot'
'Whoever needs help, he spends more time with them'
'Sometimes I don't find it (his assistance) that clear'
'It can be frustrating' (waiting for his help because he is explaining something to someone else)
'Sometimes he is idle if we don't ask for his help'
'He is doing his job; he is not nice, he is not mean, he is just there'
'I don't think there is energy and enthusiasm…I don't think I've ever seen him joke or smile'
'He is quiet…he doesn't joke but he does talk to us'
'I never felt him to be concerned about anything…he is not cold, he is not warm'
'I've seen him bored'
'Its all about marks' (i.e. not inspiration by TA)
'I'm not happy with my marks but its all relative to other people'
'He is just doing his job; just being a TA'

LOW
'He always goes around to everyone so we can rarely find him…just not enough demo for a group'
'People call him more' (i.e. students initiate interactions more)
'A little bit of hunting and you can find him'
'He doesn't come to you…we have to call him over'
'He knows which people need more help so he goes over to them more often'
'He'll spend more time with one person because they're weak at something'
'…spends time with some people more than others'
'…(sometimes) he just sits there'
'The lab is mostly on our own. If we don't know how to do it, we don't know how to do it'
'After he explains, I don't know what he is talking about so I continue sitting there cluelessly'
'We don't understand what he is talking about and he keeps on talking'
'He might just do it for us but afterwards we don't know what he did'
'When he explains he is usually telling us what to do'
'He explains everything complicated'
'He is concerned that you're on the right track but otherwise I'm not so sure'
'He is not enthusiastic…he is just not'
'He will criticize' and 'I don't find him praising us'
'No concern'
'I can't feel it' (TA's warmth)

'Oh no (inspiration). He is more like a guide along the experiment than an inspiration. We do things on our own'

'Not much (influence to attitude). I still consider anything to do with physics tedious'

'Nope' (TA stimulating interest in physics)

'No, not inspirational'

'I don't like or dislike him…its just nothing'

'He likes physics but is not enthusiastic teaching wise'

'He doesn't like to teach'

'You don't really get anything; you just do it (the lab)'

Appendix 3. Abbreviated version of the Behavior Catalogue (TABC)

I. Time And Interaction Overview

Part I of the form allowed for the recording of interaction types in tables that covered 10 minute intervals, of which the first two are shown.

SI – Student initiated interaction
TI – demo initiated interaction
NI – no interaction (>20 secs)

0 TO 10 MINUTES 10 TO 20 MINUTES

	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5		1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	
SI																SI																SI
TI																TI																TI
NI																NI																NI

II. Physical Communication

Part II of the form asked observers to rate the indicated communication variables on a 7-part Likert scale.

Volume of speech, Pace of speech, Clarity of speech, Pitch of Speech, Says um, ah, like, etc, Eye Contact, Smiles at students, Frowns at students, Physical Contact, Shows frustration, Criticizes students' Praises students' Jokes/chats with students, Uses student

names, Physical Energy (Movement), Mood, Physical Condition, OVERALL RATING.

References

Alsop, S., & Watts, M.(2003). Science education and affect. *International Journal of Science Education*, 25(9), 1043-1047.

Arreola, R. A. (1995). Developing a Comprehensive Faculty Evaluation System. Bolton, MA: Anker.

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

Frymier, A. B. (1993). The impact of teacher immediacy on students' motivation over the course of a semester. *Paper presented at the Annual Meeting of the Speech Communication Association, Miami Beach FL.*

Haskins, W. (2000). Ethos and pedagogical communication: suggestions for enhancing credibility in the classroom. *Current Issues in Education*, 3(4).
Available: http://cie.ed.asu.edu/volume3/number4/.

Hazari, Z. & Key, T.(2003). Student Evaluations of Teaching Assistants in a First Year Physics Laboratory: Is there a Grade Bias? Submitted to the Journal of Research in Science Teaching.

Henderson, D., Fisher, D., & Fraser, B. (2000). Interpersonal Behavior, Laboratory Learning Environments, and Student Outcomes in Senior Biology Classes. *Journal of Research in Science Teaching*, 37(1), 26-43.

Laukenmann, M., Bleicher, M., Fuss, S., Glaser-Zikuda, M., Mayring, P., & von Rhoneck, C. (2003) Z. An investigation of the influence of emotional factors on learning in physics instruction. *International Journal of Science Education*, 25(4), 489-507.

Lewis, R. (1998). Student evaluations: widespread and controversial. *The Scientist*, 12, 12.

MacIsaac,D. & Falconer, K. (2002). Reforming Physics Instruction via RTOP, *The Physics Teacher*, 40, 479-485.

McCroskey, J. & Richmond, V.P. (1992). Increasing teacher influence through immediacy. In V. P. Richmond and J. C. McCroskey (Eds.), *Power in the Classroom: Communication, Control, and Concern* (45, pp.200-211).

McDowell, E. E. (1993). An exploratory study of GTA's attitudes toward aspects of teaching and teaching style. *Paper presented at the Annual Meeting of the Speech Communication Association, Miami Beach FL.*

McLeod, A. (1996). Discovering and facilitating deep learning states. *The National*

Teaching and Learning Forum, 5, 1-7.

Murray, H. G. (1980). Lecturing: Classroom behaviors of social science lecturers receiving low, medium and high teacher ratings. *Ontario Universities Program for Instructional Development Newsletter,* 14, pp. 3-5.

Murray, H. G. (1977). Evaluating university teaching: a review of research. *Ontario Confederation of University Faculty Associations, Toronto.*

Nunnaly, J. (1978). *Psychometric Theory.* New York, NY: McGraw-Hill.

Pankratz, R. (1967). Verbal interaction patterns in the classrooms of selected physics teachers. In Edmund J. Amidon & John B. Hough (Eds.*), Interaction Analysis: Theory, Research, and Application.* Reading, MA: Addison-Wesley.

Sax, G. (1997). *Principles of educational and psychological measurement and evaluation.* Belmont, CA: Wadsworth.

Shavelson R., Richard J. & Webb, N. M. (1991*). Generalizability Theory: A Primer* Newbury Park, CA: Sage.

She, H-C. & Fisher, D. (2002). Teacher Communication Behavior and its Association With Students' Cognitive and Attitudinal Outcomes in Science in Taiwan. *Journal of Research in Science Teaching,* 39(1), 63-78.

SPSS base 9.0 applications guide. (1999). Chicago, IL: SPSS Inc.

Weiss, I., Pasley, J., Smith, P. S., Banilower, E., & Heck, D. (2003). *Looking Inside the Classroom: A Study of K-12 Mathematics and Science Education in the United States*. Chapel Hill, NC: Horizon Research, Inc. Available: http://www.horizon-research.com/reports/2003/insidetheclassroom/looking.php

About the authors…

A.W. Key, M.A.(Aberd), D.Phil. (Oxon)., Professor of Physics, University of Toronto. Professor Key has been a faculty member of the Department of Physics, University of Toronto since 1970. For many years he pursued experimental particle physics (Fermilab, Argonne National Laboratory, and SLAC) ; more recently his interests have been in pedagogy, psychology, teaching in higher education, and physics education research. He has recently designed a new first year laboratory for Physics Specialist students., and his current teaching includes a course for physics graduate students on Effective Communication.

John Pitre, B.Sc., M.Sc., Ph.D. (Windsor), Senior Lecturer in Physics, University of Toronto. Dr Pitre is an expert in the development of course curricula and lecture demonstrations as well as the introduction and improvement of experiments in the

undergraduate laboratories. He is the recipient of the *Ontario Confederation of University Faculty Associations (OCUFA)* Award for Outstanding Contributions to University Teaching (1994) and the *Canadian Association of Physicists* Medal for Excellence in Teaching Physics (1995).

Zahra Hazari, B.S. (Florida Atlantic), M.Sc.(Toronto). Ms Hazari is a Ph.D. candidate at Ontario Institute of Studies in Education, University of Toronto and a pre-doctoral fellow at the Harvard Smithsonian Center for Astrophysics. Her current research is a high statistics study of factors that influence success in introductory physics at universities across the United States.

---

[i] For the study TAs, the number of such times was extremely small – typically less than a dozen or so in a 2-hour period. While it made us appreciate just how very busy our TAs are, it was clear that all of them were equally so.

[ii] The design for a G study includes the sources of variability and a specification of how they are related (crossed, nested, or a combination of the two). In a crossed design, persons crossed with occasions indicates that every person was included (observed, tested, etc.) on every occasion; in a nested design, persons nested with occasions indicates that not all persons were included in every occasion (i.e. persons were observed, tested, etc., on different occasions). Our G study, with three sources of variability, 'TAs, 'occasions' and 'observers', is completely nested with observers nested within occasions and occasions nested within TAs.

[iii] On the other hand, a correlation of the students' answer to question 14 alone ('My Overall rating of my TA as a lab teacher') was not significant for the 44 TAs, whereas it was significant for the 9 study TAs. This merely confirms the unreliability of the response to a single question.

[iv] This is interesting in itself; implying that students tend to judge their TA as acting fairly insofar as they give the student a good grade; beginning students in physics, with only a high school experience behind them, are notoriously poor judges of their own ability!

[v] For the individual TA scores, the ANOVA of (SI+TI) and (TI/SI+TI) indicated that the variances of the distributions of these variables were too dissimilar for the ANOVA to be a useful test (the Levene statistic was significant at the $p<0.05$ level). The **High**, **Moderate, and Low** groupings did not suffer from this problem, and were therefore used whenever appropriate.

[vi] This analysis is weaker than the other ANOVAs since the Levene statistic, which measures the equality of variance of the different groupings, has a value of $p=0.019$ which is significant at our determined significance level of $p<0.05$.

[vii] A screen plot is a graph of the eigenvalues (representing the amount of variance explained) versus the number of factors. As more factors are added, more of the variance is accounted for. After a certain number of factors the flattening of the curve indicates that it is pointless to add more factors since not much more of the variance is being explained.