Deciding whether a teacher trainee is ready to become an effective member of the teaching profession is an expected responsibility of any teacher training institution. In many cases, more is expected. Discrimination is required toward teacher trainees in deciding who is able, who is adequate, or better or best. These expectations have necessitated institutions formulating and improving a number of methods for making such decisions. As part of the commemoration of the 10th Anniversary of the Institute of Education, this paper is being presented as a focus on the assessment of teacher trainees for accreditation in becoming part of the task force of teachers in Singapore.

There are three general areas which are considered important in the training of teachers: knowledge, attitude and performance. Typically knowledge and performance have been stressed as the principal areas in which to evaluate teacher trainees. The assessment of attitude is usually hidden in either or both of the previous two. In this paper, the areas of attitude and knowledge assessment will be touched on lightly giving the lion’s share to the assessment of performance. The sections on Attitude and Knowledge consist of a brief discussion of current thinking in these areas. The section on teacher performance entails a review of current research and thinking. Each of the three sections is followed by a brief description and evaluation of what currently is being practised at the Institute of Education (I.E.).

**Attitudes**

Although attitudes of teachers or teacher trainees have been intuitively thought to be important, they have typically been embedded into measures of teacher performance scales and, therefore, will not be treated to any great length here. However, Monroe (1981) has drawn attention to an important aspect of attitude as it relates to the teacher trainee. He contends that there is little evidence to suggest that attitudes of teacher trainees can be changed to any degree during their training. On this account, he proposes that assessment of 'potential teachers’ expectations and dispositions occurs as part of the selection procedure upon entrance to the training institute. He fails, however, to explain how evaluation of such traits at point of entry can be measured. Nevertheless, research is needed to examine the effects of attitude at point of entry upon both success during training and later in the teaching profession.

Attitudes related to teaching are considered important at I.E. for becoming a successful teacher. They are implicitly taken into account at time of entry during the personal interview as well as during the training programme itself. Provision is also made for supervisors to comment in their final evaluation on the attitude of each trainee.

**Knowledge**

An area of teacher trainee assessment that carries considerable weight in determining one's future is how much information one accumulates in course work. Attention at this point will be given to the manner of testing, leaving grading procedures to the end of the paper. Rather than cite any studies regarding this problem, a presentation of what is considered to be sound principles in achievement testing is given (cf., Gronlund, 1976; Thorndike & Hagen, 1977). The author will assume that the reader has a basic grasp of the terminology used in testing and measurement.

The manner in which knowledge is tested varies with the objectives of each course. Not only are different kinds of information required to be 'learned', but different levels of learning (Bloom, 1956) are expected as well. For example, some parts of a course curriculum will be centred more around application than simply remembering a lot of information. Where such is the case, teacher performance in the classroom is the more appropriate area to assess whether the teacher can 'apply' what s/he has been exposed to in the training course. However, many institutions continue to test these objectives with paper and pencil tests arguing that if they cannot remember the target principles they will not be able to apply them. But, if the paper and pencil test does not present a novel problem for which application of a principle is required (Anderson, 1972), the objective of Application is not being tested, therefore, decreasing construct validity. Even if a novel problem were constructed, the inference that what one writes in response is reflective of what one would do in real life is highly questionable. Since the principal issue in teacher training is what transpires in the classroom, the manner of assessment
should heavily lean toward evaluating the performance of the trainee. Course work in teacher training will, no doubt, have a certain amount of information that trainees are required to store in memory. In fear of being accused of promoting rote memorization, however, lecturers are quick to point out that their objectives in terms of the Bloom scale are higher order: Analysis, Synthesis, Evaluation. To test such abilities, however, one should shy away from ‘on the spot’ paper and pencil tests since in real life situations people are given time to display these abilities. For example, if one is given a problem in real life to analyze, and asked to synthesize a solution, one is usually allowed to go away and spend time in investigation, contemplative thinking and writing/rewriting the solution. Therefore, a method of assessment that would be more consistent to reality would be an assignment on which the trainees could spend time analyzing and synthesizing a solution. Such an assignment should be a fairly accurate representation of what the trainee’s abilities are, assuming that the work is his/her own.

As for information which lecturers believe to be important at the Knowledge and Comprehension levels, as well as the ability to display higher order levels of learning at a moment’s notice, the paper and pencil test is probably the best method. However, there are a variety of test items that could be used, from True/False to Extensive Essay questions. Whichever one is selected for use depends on what one wants to accomplish and the constraints one is working under.

The two qualities of any test which are unanimously agreed upon are reliability and validity. So-called ‘objective’ test items, such as multiple-choice, True/False, possess high reliability if constructed well, while essay items have more fluctuations in reliability especially when large numbers of people are being tested. Regarding validity, much depends on how well the test items correspond both to the number and the nature of the objectives. Objective items such as Multiple Choice can cover a considerable number of objectives and levels of learning with the exception of Synthesis. Essay items, especially if they are ‘extended’, can only cover a few objectives in one test situation. Although Essay items can test all levels of learning, they are most efficient for testing Analysis, Synthesis and Evaluation. However, very few Essay items are constructed appropriately to test these higher levels of learning. Where the situation involves small numbers of people demanding fairly sophisticated thinking, the Extended Essay item is very appropriate but, most likely, a project paper would be the principal method of assessment. On the other hand, where large numbers of students are being evaluated and a considerable number of objectives are being covered, objective test items would be most appropriate.

Current Practice at I.E.

Presently, the Diploma of Ed. programme at I.E. is divided into eight components. Seven of these components are completed at I.E involving the acquisition of knowledge and skills. The eighth component engages the trainee out in the schools.

Of the seven components involving the intake of information, four are completely assessed by assignment and the fifth, Principles and Practice of Education, is particularly assessed by assignment. The Oral Communication and Physical Education components are both assessed by progress ratings. In the light of the previous discussion on the manner of assessment, a high correspondence with objectives seems quite apparent.

Only for the Principles and Practice of Education component is an examination used. It is given at the end of the Diploma programme and consists of ‘Extended Essay’ items drawn from the various sub-components which trainees have studied over one year’s time. Whether the results of this exam are valid or reliable is seriously questioned by the author. Firstly, the Content Validity of the exam is questionable. Due to the nature of the items, i.e., ‘Extended Essay’, only a limited content area can be covered. Sampling only a limited area of content could give an inaccurate reading of all that people have learned in the various sub-components.

Second, there is a question about the ‘Construct Validity’ of the exam. The rationale behind using ‘Extended Essay’ items is that they test ‘higher levels of learning’. However, since the variation of the different exam items has been fairly predictable, and past questions accessible in a test file in the library, trainees are able to spot what is coming and memorize answers which are typically no more than a concatenation of class notes and quotations from texts. Therefore, rather than testing higher levels of learning, such exams test a trainee’s ability to commit large portions of texts to memory.

Finally, there is the question of reliability. Lecturers involved in the evaluation, are typically given a very large number of scripts to mark in relatively little time. It has not been uncommon for one lecturer to critique well over a hundred scripts. The probability of giving consistent evaluation for so many scripts is very low since extraneous variables such as mood, fatigue, and habit-focus are interfering with the assessment process.

A possible solution to these problems would
be to test each sub-component shortly after it has been completed. In so doing, each sub-component could be tested in the manner that is most appropriate. The section on the Philosophy of Education, for example, may need to set several novel Extended Essay questions in order to test trainees' ability to demonstrate logical reasoning. Some other sections which have covered a lot of ground at an introductory level may need to use a number of objective items to measure a representative sample of what has been covered. Moreover, even higher levels of learning can be measured using objective items; however, care must be taken that they are properly constructed.

I.E., at one time, used a 100 item Multiple-Choice test in conjunction with the essay exam. It was primarily used for borderline cases to determine grades but currently is no longer used. The author studied the exam and found that a number of the same items were used from year to year; in addition, some items violate prescribed item writing procedure. However, had an item bank been built up over the years of well-written items, each year's exam could be constructed from a stratified (according to subject area) random sample of items. Of course, such items should not be available in either library or student files for use by trainees. The result would be an exam with greater content and construct validity with considerable reliability.

Performance

As with any skill, the true test of a teacher's worth is in the place where they must exercise all that they have 'learned'. What s/he knows about the subject being taught and the theory of learning and instruction is necessary but not sufficient (McDonald, 1978; Monroe, 1981).

If s/he cannot transfer the knowledge of a subject or implement the principles of learning theory in the classroom effectively, then his/her knowledge is of no practical use for teaching. Therefore, it is intuitively logical that evaluating teaching ability should mainly be by way of observation of teacher performance.

The task, however, is formidable. Measuring teacher competence, whether at the pre or in-service level, has been a major challenge to those involved in teacher training programmes for over the past twenty years (Berliner, 1976; Rosenshine & Furst, 1973).

There have been two different emphases in trying to meet this challenge. Both are attempts to interrelate process with product. Process is used to define what teachers actually do in the classroom. Product involves the effect upert the pupil (e.g., academic growth, change in attitude, etc.).

One emphasis, and probably the most tradi-
of correlational studies. Rosenshine and Furst (1973) identified nine variables found to be stable enough to warrant experimental research. They related to the global categories of clarity, flexibility, enthusiasm, orientation toward task, criticism, indirectness or use of pupil ideas, allowance for learning criterion materials, use of structuring comments, and varying cognitive levels of questioning and discourse. They conclude by arguing that the next step is to subject these competencies to experimental research in order to examine their causative properties.

The stress on product as the prime criterion has led to the Competency (or Performance) Based Teacher Education (CBTE or PBTE) movement in the U.S. Popham and McNeil (Popham, 1968; Popham, 1975; McNeil & Popham, 1973) were the forerunners upon whose thinking and research the CBTE movement has been rationalised. Their strategy is to present teachers with a set of specific objectives for an unfamiliar topic with a variety of materials from which to choose. Pupils are tested both prior to and shortly after the mini-lesson is taught. Their gain scores are to be used as an index of the teacher's competency in achieving the objectives. Those teachers getting greater gain scores can be considered more competent.

Glass (1974), however, has criticised the Popham-McNeil method on several counts. First, he argues that the method is too simplistic in that it focuses mainly on pupil performances toward academic objectives. Other objectives such as attitudes and affect need to be included. Second, he maintains that it has two artificial outcomes. One outcome is that such important behaviour as maintaining classroom control would not normally be needed in the somewhat laboratory environment which the method tends to generate. The other outcome is that the procedure of presenting teachers with new materials may, in fact, take away one of the reasons for a teacher being good — i.e., having a good grasp of the material. Thirdly, Glass reviewed a number of studies which revealed that the findings are not of sufficient reliability to measure individual differences among teachers: a necessary condition for evaluating teacher differences. Glass concludes that until such reliability can be demonstrated the Popham-McNeil method "remains a laboratory technique, useful for teacher training and research, perhaps, but not up to the standards of utility, economy, and fidelity required of a teacher evaluation technique for individual diagnosis or determination of merit" (p. 17). Popham (1971) concluded himself that, at most, the procedure is only usable for groups of teachers and not for evaluating individual teacher's competencies.

Despite the above warnings, the momentum for CBTE has been on the increase, most likely accelerated by the emphasis on accountability in the U.S. Teacher education in many U.S. state programmes now have to demonstrate that they are producing competent teachers (e.g., Lukin, 1977; Stulac, et al., 1981). Unfortunately, in the mad rush to meet state requirements, a number of teacher training programmes have reverted back to using the opinion of the experts as their main criteria for selecting teacher competencies (e.g., Berrones, E., 1978; Stulac, et al., 1981).

On the other hand, there have been recent studies using pupil achievement as the criterion for selecting teacher competencies. Denton and his colleagues (Denton & Kazimi, 1982; Denton & Norris, 1972 a & b; Denton and Tooke, 1981) have carried out several studies using an expanded version of the Popham-McNeil design outlined earlier. Student teachers were observed during two 2-week periods where they were given a specific set of performance objectives for pupils to achieve. Using regression analysis, they found a fairly consistent pattern for lesson-plan development predicting pupil gain scores. Capie and his colleagues (Capie, Ellett, & Johnson, 1980; Capie, Tobin, and Bosewell, 1980) examined the validity of the Teacher Performance Assessment Instrument (TPAI) package used at the University of Georgia for teacher certification. The TPAI consists of instruments which assess teacher plans, selection of materials, classroom procedure, interpersonal skill, professional standards and perception of pupils. Following the Popham-McNeil model and using two observers they found 3 out of 17 competencies (Capie, Ellett, & Johnson, 1980) that correlated both significantly and reliably with pupil achievement: use of instructional techniques related to objectives, ability to communicate both orally and in writing, and the use of procedures involving the pupil in the instruction.

Shavelson and Dempsey (1976) point out that it is important for teacher competencies to be generalizable across situations. They argue that each teaching encounter presents different demands upon a teacher which elicit different sets of behaviours. Thus, sole observation of one teacher in one setting cannot be compared to that of another teacher in another setting even though substantial intrarater reliability may exist. What is needed, they contend, is to observe different teachers over a number of teaching encounters and measure whether these teachers maintain the same rank order on each characteristic (i.e. generalizability). They reviewed several studies which used the generalizability theory of Cronbach, Gleser, Nanda and Rajaratnam (1972) and found that global ratings were more stable than frequency counts. In addition, they found that four characteristics (teacher
presentation, positive or neutral feedback, probing and direct teacher control) were moderately stable. Six characteristics were found to be unstable: teacher questioning, negative feedback, student-centred teaching style, interpersonal behaviour, attention-getting procedure, and indirect teacher control. These findings were consistent with those of Rosenshine and Furst (1973).

The research summarized above gives hope that for student teacher assessment there are competencies which relate to achievement. The competencies which have so far been consistently associated with pupil gain scores centre around lesson preparation, communication skills of the teacher, the ability to stay on task, positive or neutral feedback and probing.

However, in line with Glass's (1974) accusation that the Popham-McNeil approach is too reductionistic, others (Erlich, & Shavelson, 1978; Gall, 1977; Pottinger, 1978) have argued that a reconceptualization of measuring teacher behaviour needs to be considered. Drawing the conclusion that a given teacher has various levels of certain competencies, based on one or two observational situations, could be misleading. They propose that a close examination of each teaching encounter be made so as to understand the constraints under which the behaviour is elicited. Rather than treat a teacher competency as a quality which will be observable in every teaching situation, competencies should be considered context specific. This is very possibly the reason that many behaviours do not show much generalizability (Erlich & Shavelson, 1978).

Other recent studies support such a proposal. Brophy (1976) reported the findings of a two year study where it was found that optimal teacher behaviour for low social economic status (SES) schools was different than for high SES schools. Effective teachers to high SES pupils tended to be more task oriented, held higher expectations and were more demanding. On the other hand, effective teachers to low SES pupils were identified as more patient, encouraging and more flexible in their teaching strategies.

In addition, an interaction between grade level and type of material with teacher behaviour has also been found. McDonald (1976) reported the findings of phase II of the Beginning Teacher Evaluation Study involving 97 teachers of 2nd and 5th graders in 45 different schools. A pre/post-test design was used measuring pupil performance in maths and reading and compared with observation of teacher behaviour. The results suggested that different instructional patterns are related to different patterns in pupil performance. For 2nd graders, an increase in reading performance was associated with combining group instruction with individual monitor-
trainee according to what they prefer. However, he notes, the favourite style of the supervisor may not be the one which the trainee can most effectively use.

Finally, there is the problem of frequency of observation. Supervisors normally observe student teachers on only several occasions. Considering the previous discussion on the evasiveness of teacher competencies (e.g., Erlich & Shavelson, 1978) one could doubt whether a reliable and valid assessment is possible based on just two or three occasions.

One other source of assessment that has been given varying degrees of importance is the cooperating teacher. In most situations, the cooperating teacher is working daily with the student teacher which could result in a more representative picture of the student teacher's ability. On the other hand, if the cooperating teacher takes on more of the role of evaluation, a strained relationship between trainee and cooperating teacher could emerge. Moreover, the cooperating teacher will have his/her own bias which most likely will enter into evaluation.

Several studies have found differences between the assessment of supervisors and cooperating teachers. Lukin's (1977) data revealed that supervisors did not regard as important the same competencies as cooperating teachers. Capie, Ellet and Johnson (1980) found that a greater number of indicators observed by cooperating teachers correlated significantly with pupil gain scores than indicators observed by supervisors. Cooperating teachers and supervisors could only agree on three competencies related to achievement. Yates' (1982) survey found that cooperating teachers were perceived by teacher trainees as having more time to evaluate and that their evaluations were considered by trainees to be more valid.

One might conclude from these findings that teacher competencies are unreliable or that cooperating teachers do not have the necessary expertise to evaluate student teachers. However, the multidimensional and context-bound situations that are characteristic of teaching experience would anticipate a difference of opinion between supervisors and cooperating teachers. Both opinions should be equally valid and important when giving appraisal of a student teacher's competence.

Persons involved in using observation methods common in student teacher assessment should receive prior training to increase reliability (cf., Borg, 1979). Both supervisors and cooperating teachers have been shown to improve their assessment skills with training (Thorlacius, 1980; Locke, 1979). By developing effective training programmes, the reliability and validity of their evaluations would increase.

Finally, McDonald (1978) has pointed out that to think a highly effective teacher can be produced in the relatively short time of a teacher training programme is unrealistic. In agreement, Locke (1979) cited a number of studies which strongly suggest that teachers do not establish their teaching style and competencies during their time in preservice training but during the first two years of actual teaching. If this is so, then evaluation of student teachers at the end of their preservice training, at most should be tentative. An ongoing evaluation over the following two years of actual teaching would probably give a more accurate picture of the competency of a teacher.

A Summary of the Review

To summarize the state of research today on what constitutes an effective teacher, one needs not be so pessimistic as to conclude that . . . "no one today really knows what the competent teacher is" (Lau, 1981, p. 135). Neither can one be so optimistic as to believe that assessing teachers' abilities is a highly reliable, well-tested enterprise. Rather, a more realistic position would be that research and theory have moved forward in the past ten years shedding more light on the subject.

First, it has become more apparent, but not surprisingly so, that the teaching situation is context specific, requiring different competencies for different types of materials under varying classroom conditions with different pupil abilities. Second, several global competencies have shown some degree of validity with classroom performance and reasonable reliability. The following are just a few: preparation and planning, flexibility, ability to communicate clearly and ability to keep pupils on task and cognitively processing information. Thirdly, there is a need for triangulation in assessing the teacher trainee; that is, several observers from different perspectives are needed to provide a fuller picture of a teacher's ability. However, rather than trying to force each observer's data into one common pot trying to say the same thing, their data represents different parts of a mosaic. A final report should reflect this mosaic. In conclusion, some doubt has been created as to whether the timing of a final evaluation is appropriate. Can it be honestly believed that a teacher's true abilities are clearly evident at the end of a relatively short teacher training programme? Or, should a final assessment be postponed until the teacher has settled in and established his/her modus operandi?
I.E. Today

The Diploma of Education programme at I.E. consists of 540 hours of course work and 260 hours of practical work in schools. The breakdown of the various components of the course-work is given on page 2 of the Handbook of Course Syllabuses: 1982-83 for the Diploma-in-Education Programme. The 260 hours of practical work is divided into a three week period for school experience and a ten week period for school practice.

The School Experience programme occurs at the end of Term I. Its main purpose is to expose trainees to real-life classrooms so that they can observe, work with, and learn from actual classroom situations. Trainees are expected to teach one or two periods a day as well as observe. During School Experience, trainees are visited once by their supervisors.

The third term of the Diploma course is entirely taken up by the School Practice programme. The objectives outlined in the Handbook for Supervisors (1) focus on the development of various skills and personal qualities to be demonstrated in the classroom. Schools are requested to allow trainees opportunity to teach about 2/3 of the 20 periods a week in their main subject and 1/3 in their second subject. They are to be assigned to two or more cooperating teachers. During the first two weeks, trainees are expected to teach 10 periods and assist and observe the cooperating teacher for another 10 periods. The teaching time increases to 15 periods a week for the following 3 weeks and then to 20 periods during the last five weeks. It should be noted, however, that there are a number of trainees who have a full 20 periods a week from the beginning of their school practice due to the shortage of staff at certain schools.

The Supervisor

Supervisors play a double role at I.E. as is common in other countries. First, they function as a facilitator to the growth and development of the trainee. Second, they have the role of principal evaluator who assesses not only the growth of the trainee's competencies but his/her final performance.

Supervisors of Diploma of Education students are usually lecturers from the student's main subject area. In some cases, where the proportion of students to lecturers in a subject area is high, lecturers from other departments act as supervisors. There are, on the average, ten trainees to one supervisor and only one supervisor for each trainee.

Each supervisor is required to visit the trainee at least four times during the ten week teaching practice. S/he is expected to be present at the beginning of the lesson and remain to observe for the entire period. Other than this, nothing else is required of the supervisor. However, there are a number of suggestions made in the Handbook for Supervisors given to every supervisor. It is recommended that s/he keeps a file on each trainee composed of comments s/he has made, comments by cooperating teacher, self-evaluations by the trainee and any additional material which may shed light on the trainee's ability. All supervisors are furnished with a Lesson Observation Pad upon which they can make comments. In addition, three instruments, (The Observation Analysis Chart, A Summary of Teaching Performances for a Complete Lesson, and the Interaction Chart) are provided with instructions on their use. There is no obligation to use any of these materials.

At the end of the sixth week, supervisors are to submit a list of trainees who might possibly get a 'A' or 'D' grade. On the average, the supervisor should have seen the trainee on at least two previous occasions. The purpose of this strategy is to allow time for a moderator to visit trainees who are considered weak or exceptionally good. At the conclusion of Teaching Practice, the supervisor submits a final evaluation of the trainee consisting of a rating of 14 competencies, comments on personal and professional qualities, and an inclusive grade ranging from A to E.

In 1980, all staff members were required to attend a workshop on Teaching Practice assessment. Since then, all new lecturers have been required to attend a briefing before they go out on supervision. In addition, each supervisor is given a handbook which explains the programme and the supervisor's role.

At the moment, there are only two requirements for being a Diploma of Ed. supervisor: being a lecturer at I.E. and holding a Diploma of Education oneself. There is no evaluation of the supervisor's own ability as a teacher nor of his/her own preferences in teaching methodology and philosophy of teaching. Additionally, there has not been any on-going evaluation concerning the quality of each supervisor's ability to provide a reliable and valid assessment of the student teacher.

The Co-operating Teacher

Cooperating teachers are selected by their principals usually on the basis of the subject matter taught. It is important that the subject taught by the cooperating teacher be that of the student teacher. They are not given any remuneration for their services.

The role of the cooperating teacher is mainly that of facilitator. Each teacher is sent a copy of the Handbook for Co-operating Teachers
published by I.E. which outlines their responsibilities. In summary, they are requested to inform trainees of pupils' individual differences, demonstrate a variety of techniques, aid in clarifying goals, etc. Presently, there is little evidence that the cooperating teacher is actively involved in fulfilling these objectives. At the end of the ten week Teaching Practice period, the cooperating teacher must file an evaluation report with the School Practice Unit (SPU) on each trainee. This report is then forwarded on to the trainee's supervisor who can take it into consideration when giving the final evaluation if s/he so wishes.

There is no formal training given to cooperating teachers other than the handbook. In 1981, workshops were given to familiarize them with the role of the cooperating teacher, but there was a poor turnout. Since then, some cooperating teachers have attended the briefings given to school principals by S.P.U. staff.

An Evaluative and Recommendation Summary

One point that quite clearly stands out in favour of the teacher training programme at I.E. is the amount of time spent in the schools. Just over 32 percent of the total time is spent either in School Experience or in Teaching Practice. In addition, trainees are given a theory-practice-theory sequence which should allow them to continually relate theory to practice.

However, with such an emphasis on the practical side of teacher training, it is imperative that the assessment of School Practice is as reliable and valid as possible. When comparing the manner in which School Practice assessment is carried out at I.E. with the findings of the previous review, several weaknesses emerge.

The primary weakness is that there are too few people involved in assessing the trainee. In most cases, only the I.E. supervisor plays the critical role of evaluator. The role of the cooperating teacher is minimal; at most, and there is no assurance that his/her evaluative comments are taken into consideration. Only in the case of an anticipated extreme score is a moderator brought in to verify the supervisor's evaluation.

A remedy which has been suggested in the literature is to make greater use of the cooperating teacher. To do so would require more training in observation and evaluation skills in order to increase reliability. Some guarantee would have to be made that top quality teachers are being selected for this responsibility. This would necessitate some sort of screening programme where teachers would be evaluated according to their own teaching capabilities.

However, the principal problem would be to increase the commitment of the cooperating teacher to the task of aiding and evaluating the trainee. The critical hindrance is the lack of time. In Singapore, there are a number of schools where the shortage of teachers forces the cooperating teacher to teach a full load while the student teacher teaches a full load in another related course. One possible solution is to bring more student teachers of the same subject area into the same school. This would release the cooperating teacher from teaching his/her course and allow time to move between classes observing, providing models, and giving help to student teachers. In addition, it would help to give the cooperating teacher some form of compensation as well.

Giving the cooperating teacher a more meaningful role would also increase the weight of his/her evaluation in the final assessment of a trainee. Rather than passing the report on to the supervisor to use if s/he wishes, it should carry equal weight and maintain its own identity. Both the cooperating teacher and the supervisor can then work together to provide a final evaluative mosaic.

The second major weakness of the I.E. Teaching Practice programme has been the lack of an ongoing evaluation and development programme. Supervisors themselves need to be evaluated on their own ability to teach as well as supervise. To insure greater uniformity, a systematic check needs to be made on whether supervisors within the same subject matter are considering the same competencies as important, as well as reliably observing them.

Moreover, more should be required of supervisors than simply attending, observing and filing a final report. Trainees themselves should be asked to evaluate their supervisors. Files that supervisors are supposed to keep on each trainee should be systematically inspected to see whether they are collecting adequate data upon which to base their evaluations.

In fact, several projects have been recently launched at I.E. which are steps forward in correcting these weaknesses. First, various departments have submitted lists of criteria on which to base their assessments. This practice should lead to greater consistency among supervisors in their evaluations. Second, the School Practice Unit is currently administering a questionnaire to trainees returning from School Practice. Trainees are asked to evaluate their supervisors in addition to other components of their ten week experience.

A third weakness has been the need for the evaluation of teacher trainees to be more context specific. The nature of the environment in which the teaching takes place should be reported. The reputation of the school and the pupils being
taught, the nature of the class being taught, and any other part of the environment that affects the teaching-learning situation need to be described. Obviously, teaching a class which has been known to have a number of trouble-makers or slow learners requires different strategies than a class with all high achievers. Or, maintaining a positive and enthusiastic attitude, in a school where full time teachers have given up and morale is low, is more difficult than when among teachers of high morale. Such descriptions should be included in the final evaluative report so that the evaluative comments can be understood in light of the constraints the trainee is working under.

Finally, evaluation at the end of the one year training programme should be tentative. At most, it should represent exactly what it is, an index of achievement and initial development toward becoming a teacher. To provide a more stable assessment of a teacher's more enduring abilities, further evaluation should be carried out somewhere in the second year of teaching. However, no evaluation is final; teaching should be continually developed and improvement constantly sought after. Unfortunately, many times, final reports during teacher training days have been hung around many a beginning teacher's neck as the albatross which always prevents him/her from being recognized as an outstanding teacher. Such a system is self-defeating.

Marking Scheme

Judging people according to their ability is as old as society itself. Typically, various symbolic systems have evolved as a means to classify the quality of one's achievements. Consequently, a variety of symbolic systems have emerged from which to choose. A controversial task of those involved in teacher training programmes is selecting one of these systems for portraying the evaluation of student teachers.

The symbolic systems which have developed vary considerably. The simplest system is distinguishing between those who pass and those who fail. Approaching the other end of the simple-complex continuum is the 15 point system that ranges from an A+ to an E-. Most programmes utilize systems that lie somewhere on this continuum.

Which of these marking systems is used depends on two factors: precision of the measuring instrument and specific values in the society the latter usually has precedence over the former. Firstly, for a marking system to be valid and reliable, it should correspond to a large degree with the measuring capabilities of the instrument(s) being used. To use a physical example, if one wants to make finer distinction in length than just short, medium and long for shoe size, one had better use a measuring device which is calibrated in centimetres. The same is true in assessing student-teachers; if fifteen categories are being used, the instrument(s) should be accurate enough to make that many distinctions. The author personally doubts if such instruments exist. On the other hand, instruments do exist that are valid and reliable enough to distinguish between those who are below average, average, above average, and possibly those who are outstanding.

The second factor, however, plays a predominant role in deciding which marking system is to be used. During the history of any society a system of values develop which determine who receives rewards and for what. Within a short time, societies become the slaves of such reward systems making them self-perpetuating. Later, when the society realizes that the system is no longer meeting its purpose, or that it was the wrong one in the first place, tradition as a relentless taskmaster whips the society into compliance with the old system.

Such an enslavement seems to have occurred in the assessment of student teaching. Even though it appears that the instrumentation is not suited for making fine distinctions, programmes continue to use marking systems that require them. In defence the reason usually given is that people in the Ministry of Education* and principals of schools are familiar with the old procedures: to change would confuse things. More importantly, however, the salary and promotion schemes are tied to the marking system, and to tamper with the 'Honours' system would almost be sacrilegious.

An additional factor which must be taken into consideration is the individual variation on the trait being assessed. That is, the degree to which the group is heterogeneous or homogeneous should be reflected in the final assessment. If the group consists of a representative sample of the total population, one would expect a distribution approaching the Normal Curve. On the other hand, if any factor is in operation, which samples from only part of the population, such as selection criteria, the group would be quite homogeneous. Making a large number of discriminations in such an 'alike' group would be extremely error prone.

Most teacher training programmes employ some form of selection procedure to identify prospective teachers. Furthermore, Diploma of Education programmes require applicants to have previously completed university. These two factors alone should not only result in a fairly homogeneous group being admitted into the
programme, but would select people from the top end of the total population distribution.

For a marking system to have acceptable validity, it must reflect, to a considerable degree, the nature of reality. From the above argument, one would expect whatever system being used to make very few distinctions in assessing teacher trainees. Moreover, such distinctions should represent a high level of successfullness. Unfortunately, some people continue to force fairly homogeneous groups of teacher trainees to fit the Normal Curve model. The results, of course, are invalid and, therefore, misleading.

Uniformity of Marks Across Subject

Recently at I.E., the question has arisen regarding the uniformity of grades across subject areas. Theoretically, if one were to hold the position that grades should be uniform across subjects, one would have to satisfy two constraints. The first one is that the objectives and related criteria for demonstrating the achievement of such objectives are similar between subject areas. For example, a subject area that emphasises communication skills has different goals than one stressing principles of learning. The ubiquitous university professor who is very knowledgeable and yet cannot teach his/her subject is well known to many. Seldom, can one find subject areas that agree as to what should be emphasized. Each area has its own somewhat isomorphic set of objectives.

The second constraint for uniformity of marks is that the distribution of individual differences on traits required to achieve the objectives do not differ between subject areas. This constraint involves selection criteria at point of entry into the training programme or a condition for entering a special subject area. To illustrate, if a subject involves communication skills, it is likely that the average university graduate could not master it, then the expected distribution of marks should be fairly homogeneous with many scoring in the top categories. If additional selection criteria are used which relate to a particular component, the expected distribution would be even more homogeneous with even more scoring in the top category than characteristic of other components. Failures would not be expected except for the few who are not properly motivated or as a result of some other extraneous factor.

At the same time, however, teaching involves more than just being able to respond on a 'paper-and-pencil' test. A teacher must be able to communicate to a group of pupils effectively. S/he needs to be perceptive, flexible and creative. S/he must be able to keep order when pupils get restless. These and many other factors are not reflected in previous performance on academic subjects and will, no doubt, play a role in making distinctions among trainees that previously have not been made.

Distribution of Marks

A more important question is whether the distribution of marks is consistent with what is known about each subject area and the trainees involved. One factor which must be kept in mind is that the nature of the subject matter interacts with the ability of the trainee. Given a sample of above average students in physics, for example, who have taken a course in astrophysics, one would expect to have a small number of those who perform outstandingly, a large number who do quite well, and a small group who do acceptable work. There would be no expected failures; although, the possibility would still exist due to such factors as motivation, illness, etc. If, however, the same group of students were given a general course in physics, one would expect the majority, if not all, to perform in an outstanding manner. Being able to distinguish between them would be practically impossible.

The same principle is true for estimating the expected distribution of marks for various components of a teacher training programme. Those involved in administrating each component need to analyze the difficulty level in relation to the quality of trainee. If the nature of the component is of moderate difficulty, i.e., there is no reason the average university graduate could not master it, then the expected distribution of marks should be fairly homogeneous with many scoring in the top categories. If additional selection criteria are used which relate to a particular component, the expected distribution would be even more homogeneous with even more scoring in the top category than characteristic of other components. Failures would not be expected except for the few who are not properly motivated or as a result of some other extraneous factor.

Sources of Deviant Distributions

A failure to obtain the distribution mentioned above would be a result of either one, or a combination, of three factors. First, the selection programme is not functioning efficiently. That is, trainees are being chosen who are not, in fact, of
the quality the selection criteria predict. The second possibility is that given that the quality of trainees is as it should be, the training programme itself could be ineffective. Few would want to even consider this one, although it is a real possibility. Thirdly, given an effective selection and training programme, something could be amiss in the assessment programme. Most likely, highly subjective criteria which usually do not relate to the objectives of the curriculum — e.g., handwriting style — have been introduced. This is an old 'trick-of-the-trade' used to spread out a homogeneous distribution in order to justify one's marking scheme; however, it reduces, the validity of the assessment results.

On the other hand, obtaining distribution with a frequent number of high marks has been interpreted by many as a result of having low standards. Unfortunately, rather than stand one's ground on sound principles of measurement, many a programme has been intimidated into imposing marking schemes that make artificial discriminations between individual trainees.

A major culprit at fault is the 'quote system'. That is, somewhere in the history of various programmes someone has decided that only a certain percentage of A's, B's, C's, etc. are to be given, usually based on an over-generalization of the Normal Curve. To perpetuate this practice, reward systems have accrued which honour one or two of the outstanding candidates. Once such practices become tradition, they are extremely difficult to change.

In summary, the argument is made that the basis for a marking scheme of any teacher training programme should be the estimated interaction between the curriculum and the quality of trainee. Where trainees are regarded as of high quality and nothing exceptionally difficult characterizes the curriculum, fairly homogeneous and high levels of performance should be expected. Distributions of marks are indicators of whether a programme is on target. Deviations from expected distributions should be a warning that something is not functioning correctly. The possible sources of difficulty are the selection procedures, the training programme, and/or the manner in which the trainees are being assessed. Such factors as tradition should be removed from the marking procedures as quickly as a society will allow.

### Current Practice at I.E.

Entry requirements for the Diploma of Education programme at I.E. are several. The major one is that all applicants are required to have a degree from the National University of Singapore or from other universities acceptable to the Public Service Commission of Singapore. If applicants have previously done well on the English component of their O-levels, they are exempted from taking the English Proficiency Test: the second requirement. If the applicant is applying to the Chinese medium component, s/he must meet an additional set of fairly stringent requirements that are made by the Chinese Department. Finally, each applicant is personally interviewed in order to insure a proper attitude and to test his/her ability to communicate. Applicants for specialized areas such as science and maths must have majored in these subjects at university.

### Table 1

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>English</td>
</tr>
<tr>
<td>1981: 185 (83.7)</td>
</tr>
<tr>
<td>1982: 290 (88.7)</td>
</tr>
<tr>
<td>Chinese</td>
</tr>
<tr>
<td>1981: 156 (100)</td>
</tr>
<tr>
<td>1982: 95 (100)</td>
</tr>
</tbody>
</table>

Note: Numbers in brackets are percentages.

### Table 2

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Bachelor's</td>
</tr>
<tr>
<td>1981</td>
</tr>
<tr>
<td>English: 3 (1.4)</td>
</tr>
<tr>
<td>Chinese: 0 (12.2)</td>
</tr>
</tbody>
</table>

| 1982                                   |
| English: 7 (2.1)                       |
| Chinese: 1 (1.1)                       |

Note: Numbers in brackets are percentages.
Table 2 provides a breakdown of the previous credentials of those accepted into the Diploma programme over the years 1981 and 1982. One can see that for the English group over one third either possess an Honours degree or a Master's. For the Chinese group the proportion was much lower. However, in 1982 there was a decrease in the number of Pass degree holders in the Chinese group causing the proportion of Honours degree holders to increase.

Taking all of the above information into account, one would expect a distribution of marks that reflected a considerably high level of performance. This is especially true for the English group who have over 30% trainees who had previously obtained an Honours degree. However, other factors previously mentioned such as ability to stand before a group of pupils, motivation, etc. will introduce some variation into the distribution and increase the likelihood of making some distinctions.

Rather than report in this paper how well different schools and departments at I.E. have produced distributions to fit these expectations, the author leaves that task to them. The proposal presented here is that departments and schools do not compare their distribution of marks with one another. Rather, they should compare the distribution of marks with the above expectations. If they find considerable deviation, they should investigate either the quality of their training components and/or the reliability and validity of their assessment programme.

References


