Title of Unit: Assessment

Student: ——————————

Unit tutor: Dr. J Lowe

Title of assignment: **Distinguish between the terms "criterion-referenced assessment" and "norm-referenced assessment". Describe and explain, from within your own educational practice, specific situations in which each type of assessment could be justified. Are these two approaches to assessment always mutually exclusive?**

Introduction

This assignment aims to compare and contrast the key elements in not only "criterion referenced" and "norm-referenced" assessment but also "ipsative-referenced" assessment. The importance of these forms of testing will be studied including an analysis on how T————— School in ————— uses these techniques.

T————— is a private day school for 600 boys (12 to 18) and girls (16-18). It is one of the most academic schools in the United Kingdom, with 20-25% of students gaining Oxbridge places. In fact the headmaster proudly states on the school web site that in each of the last six years 85% of A level entries

have been awarded grades A-C. In August 2001 the school was placed ⸺ place in the *Guardians* league table for successful schools (based on A level results.) Assessment is an important component within ⸻ and the school has recently reviewed its system of grading. The essay looks at the schools approach towards assessment, and analyses if the three methods are mutually exclusive.

The three forms of assessment

Ipsative-referenced assessment is based on the individual performance of a student. The performances of other students are not taken into consideration. Therefore the assessment looks at how much the student has improved from an earlier date. Athletics training often uses ipsative assessment, concentrating on the improvement of an athlete using his or her own individual previous scores as a benchmark.

The photograph overleaf illustrates that the athlete has successfully improved his lifting ability by 2.5kg.
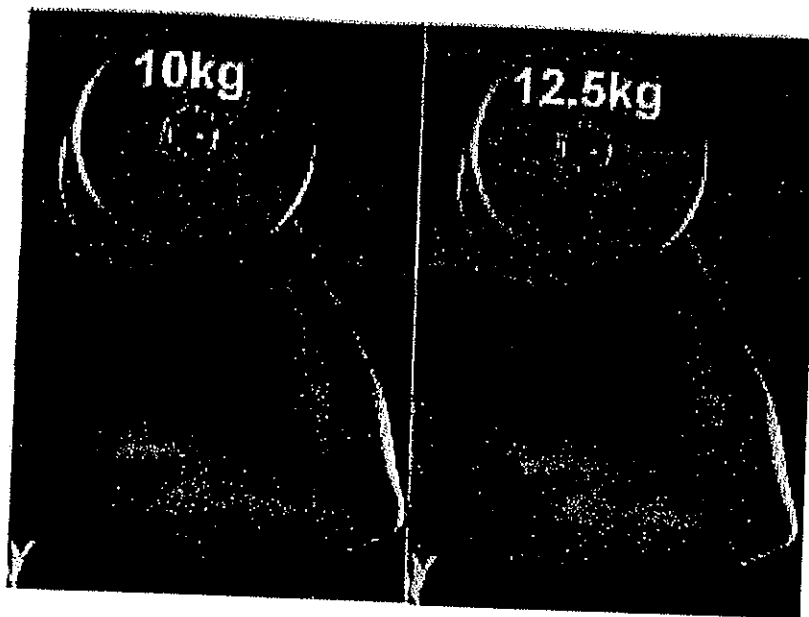
The emphasis is not on how much the athlete has lifted, but on the improvement in the lift. The performance of other athletes is not taken into consideration.
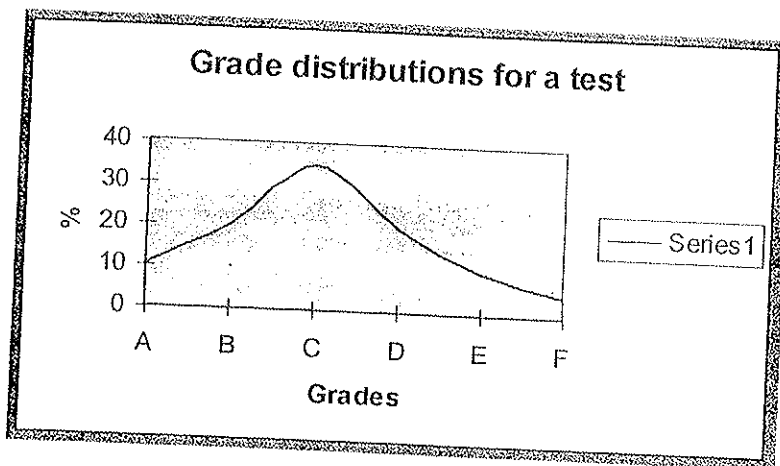
Norm and Criterion-referenced assessment

According to Dylan Wiliam (1997) much of the history of educational assessment has involved the comparison of an individual with a group, or "norm". The increased criticism of this form assessment over the last forty years has led to the development of criterion-referenced assessments where a student's performance is related some well defined objective.

Norm-referenced assessment is based on a comparison between the relative performances of students. This is achieved by either comparing the performance of individual students within a group, or by comparing the performance of an individual with a "norm" defined by the performance of an appropriate group. Such assessment may simply involve ranking the

students, or may involve scaling their marks or grades so that they fall on a standard distribution of some sort. It may also involve establishing arbitrary pass rates, mark distributions or grade distributions, e.g. only 10% of students will be awarded 'A' grades, 20% 'B' grades, 35% 'C' grades, and so on.

The graph on the below helps to illustrate this. The graph illustrates that 10% of students will gain an A grade. These students gained the A grade because they gained a better score than the 90% of other students. It does not necessarily indicate how well the students have done, as they gained their grade on a relative basis.

Grade distributions for a test

Norm-referenced assessment has a long tradition of use in national and professional examinations. In the United States for example the Scholastic Aptitude Test (SAT) is widely used as an entrance qualification. Until recently the performance of each student was compared with a group of students

from the northeast United States, who took the test in 1941. Therefore a score of 500 relates to the average performance of the "norm" group. 700 would equate to the top 2% of 1941 students. According to Dylan Wiliam (1996) there are serious weaknesses with this approach: *"The disadvantage with this approach is that the meaning that we can give to the scores in the test depends entirely on how the "reference group" is defined. In the case of the US SAT, the reference group is now highly unrepresentative of the college-bound population."*

The major reason for using norm-referenced tests, according to Linda Bond (1996) is to classify students. *"NRTs are designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers.* "She continues by stating that schools may want to classify students in so that they can be streamed to help teachers group students of different ability levels of reading or mathematics instructional groups.

Criterion-referenced assessment is based on what the student can or cannot do, regarding external criteria specified in advance. This form of assessment is often carried out on a pass/fail basis. Vocational training often uses criterion-referenced assessment. The classic example of this is the driving test. A student will pass or fail depending on if they can complete the set tasks. If the student cannot complete a three-point turn then they will fail.

The diagram below illustrates an exam mark sheet, where the examiner has observed that the student has successfully completed task A to D. The success of the student is based on these tasks, and is not related to the ability of other students.

```
┌─────────────────────┐
│  •  Can do A  ☑     │
│  •  Can do B  ☑     │
│  •  Can do C  ☑     │
│  •  Can do D  ☑     │
└─────────────────────┘
```

According to J Lowe (2000) criterion-referenced forms of assessment have obvious strengths. *"The criterion-referenced approach has immediate appeal in that it gives us information about actual learning outcomes of the system being monitored. Unlike a norm-referenced model it tells us what learners know or can do."* D Baume (1999, p19)) adds to this, recognising a number of advantages for criterion-referenced assessment. He states that it clearly identifies what is valued in a curriculum, teachers know exactly what to teach, level descriptors (as used in GCSE and A levels) make it clear to students what they need to learn, and help provide feedback identifying what students have achieved. Baume however also recognises weaknesses with criterion-referencing. The main problem stems from the complexities that arise this form of assessment.

During the 1980's the British government changed the school exam structure, by replacing O levels with the GCSE. One of the many aims of this change was to convert from a norm-referenced assessment to one that is criterion based.

Caroline Gipps states that another major reason for introducing the GCSE was due to concerns over comparability between exam boards. The intention was to create a single, consistent award with clearly defined goals. Yet another reason however was due to *"an attempt to boost standards: Sir Keith Joseph's aim was to get 80%-90% of 16 year olds up to the level previously deemed to be average. On norm-referenced tests there is no point in trying to get every pupil to achieve an average or above-average score since, by definition, these tests are designed to have half the population scoring above and half below the mean. With criterion-referenced assessment in theory everyone can achieve the top grade."* Since then it has been possible for the number of students with A grades to grow. This does not necessarily mean that the exams are necessarily getting easier.

There are however difficulties when introducing a criterion-referenced assessment. There are few problems for simple assessments such as indicating if a student can swim 25 metres, but within complex subject areas the difficulties mount. Requirements become abstract and unreliable, and the marking schemes complex.

GCSE's, the AS and A2 levels have attempted to introduce criterion-referenced marking schemes, but have found it difficult, according to Baume (1999) to *"develop statements of learning outcomes, criteria for awards at different levels, and formulations of thresholds. For example, there are problems of conceptualisation. What exactly is involved in, for example in "critical thinking"?"* Attempts to produce criteria to overcome this often lead to long-winded, complex documents, whilst cutting down and simplifying them may lead to both student and examiner not understanding what is required. The table on the next page illustrates the assessment criteria for the AQA GCSE Geography coursework paper.

Assessment Criteria

Criteria for assessing coursework component.

	Level 1 Marks 1-2	Level 2 Marks 3-4	Level 3 Marks 5-6
Applied Understanding	The candidate locates the study area in a basic manner and through brief description, demonstrates some understanding of the ideas and concepts involved and can apply them in a simple manner to the geographical topic. Uses a limited range of geographical terminology.	The candidate locates the study area and demonstrates through description and explanation an understanding of the ideas and concepts involved and can apply them to the geographical topic. Uses a range of geographical terminology.	The candidate locates the study area in detail and demonstrates through description and explanation a thorough understanding of the ideas, concept and processes involved and can apply them constructively to the geographical topic. Uses a wide range of geographical terminology.
Methodology	The candidate identifies a question or issue and lists the methods used in obtaining the information. Selection, observation, collection and recording uses a limited range of basic techniques.	The candidate identifies a question or issue, the sequence of investigation and describes the methods used in obtaining the information. Selection, observation, collection and recording uses a range of appropriate techniques. The work is organised and planned and shows some evidence of the development of tasks.	The candidate identifies a question or issue, explains why that particular question or issue was chosen. The candidate describes the sequence of investigation, the methods used in obtaining the information and explains why the methods selected are relevant to their investigation. Selection, observation, collection and recording uses a comprehensive range of appropriate techniques. The work is well organised, planned and shows evidence of originality and initiative by the candidate.

Data Presentation	The candidate uses a limited range of basic techniques, some of which are ICT based, to present the information and express simple ideas with some degree of accuracy.	The candidate uses accurately a range of techniques, some of which are ICT based, to present and develop the information; and express ideas with considerable accuracy in the use of English.	The candidate uses accurately a range of more complex techniques, some of which is ICT based, to present and develop the information appropriate to their investigation; and express ideas in a clear, fluent and logical form using precise and accurate English.
Data Interpretation	The candidate gives a brief description of the results and/or suggests basic reasons for the results.	The candidate makes valid statements about the results. Attempts are made to analyse the results. Conclusions are drawn that relate to the original purpose of the enquiry.	The candidate demonstrates links through a detailed analysis of the material. In referring specifically to the data valid conclusions are drawn that relate to the original purpose of the enquiry.
Evaluation	The candidate briefly describes how the enquiry process can be improved by questioning the reliability of the methods used to collect the data.	The candidate describes how the enquiry process can be improved by questioning the reliability of the methods used to collect the data and/or the accuracy of the results.	The candidate describes how the enquiry process can be improved by questioning how the reliability of the methods used to collect the data have affected the accuracy of results and the validity of conclusions.

The AQA GCSE marking criteria illustrates that students will gain a level 3, or high grade, if within the Data Presentation section "*the candidate uses accurately a range of more complex techniques.*" The difficulty for both examiner and candidate is defining what "*more complex*" means compared to the level 2 specification. The same problem arises when assessing if the candidates work should fit into level 1 or 2.

Nicholas Pyke, writing in the Times Educational Supplement in 1998 recognised this problem with the A levels, quoting Dr. Mike Cresswell of the Associated Examining Board. The problem of abstract requirements within marking schemes have led according to Cresswell for examiners marks to vary wildly between years, resulting in modification procedures that bring the results into line with expected grades. Pyke states: "*True "criterion referencing" of the type sought by ministers is unfeasible on a mass basis. It would mean asking the same questions every year.*" This has led Dylan Wiliam (1997) to state that "*any criterion-referenced is underpinned by a set of norm-referenced assumptions, because the assessments are used in social settings.*" The GCSE and GCE exams therefore are a combination of both criterion and norm-referenced assessment.

T———— School

As mentioned earlier in this essay T———— is a private secondary school, noted for its academic success. It can be seen that all three assessment

techniques are used in the school, but the importance of each varies between different school activities and areas.

Entrance Examinations

Students wanting to "win" a place at T———— sit exams, set by the school when they are either ten or twelve years of age. 72 students gain entry at ten years of age, whilst only 25 gain entry at twelve. These exams are norm-referenced, and the students directly compete against other students who sit the exams on the day. The students sit exams in verbal reasoning, non-verbal reasoning, Maths, and English. They are also interviewed, and are asked to bring along pieces of work, such as art to present. About 200 students sit the entrance exam. These exams are norm-referenced, and the students directly compete against other students who sit the exams on the day. However well each student performs, their success in gaining entrance is dependent on the other students, and if 72 students gain a better score than themselves they will not be offered a place. Norm-referencing therefore plays a fundamental role in the school. Many potential students may feel unlucky that they failed to gain entrance due to the many strong candidates of that particular year, whilst others lucky enough to gain a place may not realise that they only succeeded due to a particularly weak group of candidates for the year they applied. Although this seems a weakness in the method of entry, the school is not a particularly wealthy school, charges "moderate" fees and cannot afford not to fill classrooms even if they are presented with a group of relatively poor students.

External Examinations

Students within T———. like the great majority of students within the UK sit both the GCSE and GCE As and A2 level exams. These external exams play an important part within the school and much of the student's life is devoted towards preparing them for these exams. These exams are criterion-referenced exams as mentioned earlier in this essay. Within Years 10, 11, 12 and 13 much class work is dedicated to preparing for these exams. I teach within the Geography Department, and we set regular end of unit tests (about once every 5 weeks) as well as exam technique lessons and mock exams. The pupils work is graded as a GCSE/GCE paper. Therefore many of the internal marks and grades within those year groups are criterion-referenced. A student will gain an "A" grade for a piece of work if he has reached a certain level of achievement, and is not dependent upon the success or failure of his colleagues.

Internal Assessment

In November 2001 the school introduced a marking and assessment policy, intended partially to formalise assessment procedures. The paper has helped in illustrating how the school assesses its pupils. The paper advises the staff to assess their pupils in a variety of ways, with the following aims in mind:

1. Motivational - Marking should praise good work and provide realistic goals for students.

2. Diagnostic role - Marking should demonstrate where things are going wrong for an individual.

3. Prescriptive - Marking should provide pupils with strategies for improvement.

4. Summative - Marking should record the achievement of pupils over a period of time so that accurate grades can be compiled.

5. Informative - Marking should provide information about the effectiveness in the teaching of a certain topic.

6. Differentiation - This will assist in the setting and grouping of pupils

7. Self assessment - Pupils should be encouraged to make honest judgements about the quality of their achievements.

8. Contextual - Marking and assessment should occur in the context of relevant baselines (eg MidYIS, GCSE grades). Where an individuals progress is better than expected, suitable praise and encouragement should be given.

The broad aims of the assessment policy as outlined above illustrate that the school recognises that all three assessment techniques are important. Point 6 for instance illustrates the importance of norm-referenced assessment within the school marking policy. Points 1,2 and 3 however encourage teachers to use ipsative assessment with goals being set for students and praise being given for good work.

The marking and assessment policy has helped the school not only with its Internal grading throughout each academic term but also with report writing.

A full report is written on each pupil at the end of each term. The report includes both an assessment and diligence grade, a written comment by each subject teacher and a tutors report at the end. A letter is sent to parents each year explaining the grading system. A section of this letter is shown below:

REPORTS: ADVICE TO PARENTS

Some explanation of the grading system may be helpful:

1) Please remember that T————. is a highly "academic" school in terms of the pupils it takes in; our gradings reflect that, and are **not** based on national averages.
2) The grades relate broadly to standards within a yeargroup, and to our experience of yeargroups over a longish period. The grades are **not** related to GCSE, AS, A2 or A level standards or predictions.

Gradings:

Attainment - (expressed as a reflection of the yeargroup):   Diligence:

1*	Exceptional (5%)	)
1	Very Good	) c 70%
2	Good	)
3	Moderate	
4	Weak	

A - Outstanding
B - Very Good
C - Good
D - Unsatisfactory
E - Poor

Staff have been asked to take the following components into account in assessing diligence:
- Whether or not homework is completed in full, well presented and on time
- Whether or not the pupil is well prepared/organised for the lesson
- The degree of positive response and/interest in class
- The degree of readiness to improve

The letter above illustrates that norm, criterion, and ipsative assessment techniques are all being used. The grading system for reports however does show that norm referencing is a very important element. Only 5% of each class are expected to receive the highest grade, and 70% will receive the next

highest. The students therefore are in direct competition. The first sentence confirms this by stating that grades are not based on national averages, but "*grades relate broadly to standards within a year-group.*"

Many teachers find this grading system difficult to understand; on the one hand teachers are asked to grade students in certain year groups in line with national external exams (i.e. a criterion-referencing system), but a norm-referencing system is required for the end of term reports. This is a "bone of contention" for many teachers who feel that this confuses both parents and pupils, and leads to many students being left disappointed. A student who has gained "A" grades throughout the term, but ends up with a "moderate" 3 might feel a little despondent. One change that has occurred within the last three years for this very reason has been to remove the "class position" section. It was recognised that this form of norm-referencing, as Dylan Wiliam (1996) states "*does not tell us what an individual can or cannot do.......which many people see as potentially unfair.*"

Ipsative grading is also evident within the diligence marks, where high grades are given the students who demonstrate a "readiness to improve."

An important element within the report assessment is obviously the written comments by both the teaching staff and pastoral tutors. An example of a subject report sheet is shown on the next page:

Report on M ——————

2 A

| Attainment grades | Absent   sessions |

1* – Exceptional   1 – Very Good   2 – Good   3 – Moderate   4 – Weak

Diligence grades

A – Outstanding   B – Very Good   C – Good   D – Unsatisfactory   E – Poor

English   Att [2] Dil [B]

M——— is making steady progress and gaining confidence in this subject. His book review skills are developing well. Punctuation and structuring sentences still need attention and should improve with more practice. He contributes well in class.

French (Set 2)   Att [2] Dil [B]

M——— has achieved consistently good results in this subject. More attention could be paid to writing skills, and more contribution in class would be appreciated. This being said, however, he has demonstrated considerable progress this term. Keep up the good work!

Latin   Att [2] Dil [B]

A conscientious and well-motivated student who contributes willingly and accurately in class

History   Att [2] Dil [B]

His work is neat and well organised. In general he shows a good grasp, though there is some slight lack of depth in more analytical work.

Geography   Att [1] Dil [A]

The report above shows clearly that ipsative-referenced assessment takes place within the comments made by subject teachers. The French teacher for instance describes how the pupil, M——— has demonstrated "considerable progress" this last term.

Value Added Testing

An important development within schools over the last decade has been the implementation of value-added monitoring systems, which helps show the progress students make within the school. Tom Roper has recognised the

importance of this development concerning the reporting of national test results in league tables, stating that *"concern is growing that league tables based upon raw scores do not present a fair basis for comparison. Increasingly education is looking at the idea of value-added as basis for fairer comparison between schools."*

T————— uses the YELLIS and MidYIS value-added monitoring systems for a number of reasons. These include providing the school with an indication of the potential performance of individual students, and monitoring their development and weaknesses.

The YELLIS website homepage also states that "*the value added approach allows you to make a fair comparison of the progress made by your students with that made by other all other students participating in the YELLIS project. To make these comparisons each student needs to be measured against a common baseline at the beginning of the YELLIS project with GCSE exams providing an "end measure".*" This is an important statement since it illustrates that this form of monitoring can be a form of norm-referencing ("*a comparison of progress*") though criterion-referencing tests are also used ("*with GCSE examinations providing an "end measure".*") The fact that schools use these to illustrate the development of individual students and monitor problem areas suggests an ipsative-referenced form of assessment.

Conclusion

This essay has distinguished between ipsative, norm and criterion-referenced assessment techniques. It has illustrated the strengths and limitations of each, and has shown how criterion-referenced assessment has developed since the 1960's due to a growing concern over the limitations of norm-referenced testing.

The essay has also demonstrated how these three very different assessment techniques are all important within the life of T——— School. The teachers use all three techniques when testing, grading and writing reports, whilst external exams use both norm and criterion-referencing techniques.

Added to this Dylan Wiliam (1996) recognises that "*if you scratch under the surface of any criterion-referenced assessment you will find a norm-referenced set of assumptions lurking underneath.*" He uses the driving test example, requiring a candidate to turn the car around using forward and reverse gears. This is commonly referred to as a "three point turn." The candidate will pass if they use a four or five point turn, but would fail if they needed 43 changes of gear, even though they have met the given criteria. Wiliam (1996) also recognises the importance of ipsative assessment within the classroom. It is clear that teachers make hundreds of assessments each day. Most of these according to Wiliam are in fact ipsative; "*When a student asks for help, she makes an assessment of what kind of response would be appropriate for that individual, on that occasion. All these assessments are*

ipsative, and in many ways they are the purist kind of assessment because they are driven only by the need to find out where the student is, so as to help them learn more effectively."

What is clear from this essay is that all three forms of assessment are mutually inclusive within schools such as T————— School —————————. The assessment might be part of an informal situation as described in the previous paragraph, or it might involve the formal entrance examinations for example. With each case both criterion and norm-referenced assessment play an important role within the school.

References

Baume David: *A briefing on key concepts. Formative and summative, criterion and norm-referenced assessment.* Assessment Series n.7. Generic Centre Learning and Teaching support Network, 1999

Bond Linda: *Practical Assessment.* Research and Evaluation Periodical, 1996

Lowe, John. *Assessment and educational quality: implications for international schools.* International Schools and International Education (Hayden and Thompson) 2000. Kogan Page

Gipps Caroline and Stobart Gordon: *Developments in GCSE.* University of London. Institute of Education.

Pyke Nicholas: *Comparisons are Tedious.* Times Educational Supplement. October 23rd 1998 p5

Roper Tom: *The National Curriculum and its Assessment Lecture.* University of Leeds School of Education.

Wiliam Dylan: Standards in Examinations: A matter of trust? The Curriculum Journal Vol 7 n3 Autumn 1996

Wiliam Dylian: *Construct-referenced assessment of authentic tasks: alternatives to norms and criteria.* Paper presented at the 7ᵗʰ Conference of the European Association for Research in Learning and Instruction, Athens, Greece, August 1997

YELLIS internet homepage: http://cem.dur.ac.uk/Yellis/default.asp

Bibliography

Garden Sue: *You've all done very well,* The Source Public Management Journal, 2001

Massey A J: *Criterion-related Test Development and National Test Standards.* Assessment in Education, Vol2 N2, 1995.

Wiliam Dylan: *Some technical issues in assessment: a user's guide.* British Journal of Curriculum and Assessment Vol2 N 3 pp11-20, 1992