

# *Setting Standards*

**By**

**Mbithi wa Kivilu**

**UMALUSI-Conference Held at CSIR Pretoria**

**September 21<sup>st</sup> 2004**



# Components of standards setting process

- ❖ **Academic content standards-** reflect the ideas, skills, and knowledge in each discipline that are important enough for everyone to learn.
- ❖ **Assessment system-** Classroom assessment, public examinations & National assessment programmes (e.g. Systemic assessment)
- ❖ **Performance standards (sometimes called indicators)-**define “excellence” and how good is “good enough” It defines required levels of mastery.
- ❖ **Proficiency standards (levels)-** assign value to example of student work expected at certain developmental levels.
- ❖ **Effective implementation/Accountability Strategies**



# Qualities of any good set of standards

- ❖ **Rigorous-demanding, set high expectations**
- ❖ **Intelligible-Clear and understandable**
- ❖ **Measurable-Students can produce work products capable of being measured**
- ❖ **Specific-Are neither so broad as to be vague nor so narrow as to be trivial**
- ❖ **Comprehensive-Adequately cover all essential areas of a subject.**



# Cont.

- ❖ **Academic-Guard** against those who view standards as vehicles to change students' social or political views or to modify behaviour. Should remain focused on all-important academic concerns.
- ❖ **Balanced-** Between skill and knowledge(e.g Library book)
- ❖ **Manageable-** Offer a road map for teachers (offer guidance to teachers in developing curriculum activities, classroom materials and instructional methods). Give teacher latitude
- ❖ **Cumulative-Learning** is a building process. Early grades should provide foundation for higher learning in succeeding grades.



# Performance standards

- ❖ **To test developer and psychometricians, performance standard usually refer to a point on a test score scale that separates one level of achievement (e.g. pass) from another (e.g. fail) identified through a technically sound process.**
- ❖ **To educators performance standard often means a description of what a student knows and can do to demonstrate proficiency on a content standard or cluster of standards.**
- ❖ **To others, the term indicates examples of student work that illustrates world-class performance.**



# Steps in crafting a set of performance standards

- ❖ **Set the number of performance levels**
  - ❖ **A single level would limit discrimination on student knowledge**
  - ❖ **Too many levels would make it difficult to discriminate between levels (recommend 3-5 levels)**
- ❖ **Name the levels e.g advanced, proficient and basic**
- ❖ **Provide descriptions of the content and quality of performance at each level**
- ❖ **Test items are developed and administered**
- ❖ **Cut scores (or levels) are decided**
- ❖ **Student work samples are provided.**



# Statistical methods of setting standards

- ❖ **Difficulty of setting defensible standards on tests.**
- ❖ **There is no agreement on a best method, although some procedures are far more popular than others.**
- ❖ **Jaeger (1976:2) argues that all standard-setting is judgmental. No amount of data collection, data analysis and model building can replace the ultimate judgmental act of deciding which performances are meritorious or acceptable and which are unacceptable or inadequate.**
- ❖ **All that varies is the proximity of the judgment of a standard to the original performance.**



# Cont.

**Recognition of standard setting as a judgmental process brings to the fore an array of important and interrelated questions.**

- ❖ Who should make judgments?**
- ❖ How should judgment be elicited?**
- ❖ Should judgment be based on information about tests, items, testing applications, examinee performance or a combination of these factors?**





# Standard-setting models

- ❖ **Test-centered models and**
- ❖ **Examinee-centered models.**
- ❖ **A seemingly exhaustive listing of standard-setting models can be found in Berk (1986).**
- ❖ **South Africa (Umalusi) uses a combination of these models in setting standards.**



# Angoff's procedure

- ❖ Requires that each of a sample of examiners examine each item on a test and estimate the *probability* that the “minimally acceptable” person would answer each item correctly.
- ❖ The examiners would think of a number of minimally acceptable persons, instead of only one such person and would estimate the proportion of minimally acceptable persons who would answer each item correctly. The sum of these probabilities, or proportions, would then represent the minimally acceptable score (assuming each item has a score of one).



# Angoff's procedure (Cont.)

- ❖ **Identify an appropriate population of examiners and**
- ❖ **Select a representative sample from that population.**
- ❖ **Each examiner must then be given some guidance on the conceptualization of minimally competent examinee.**
- ❖ **The examiners can identify the level of knowledge and skills that, in their view, defines the distinction between students who are acceptably competent and those who are not.**



# Angoff's procedure (Cont.)

- ❖ **If agreement on borderline knowledge and skills can be reached, definitions are written together with examples of student performance, that are above and below the borderline.**
- ❖ **Once examiners understand their task clearly they consider each item on a test individually and decide for that item the probability that a minimally competent examinee would be able to answer it correctly.**



# Iterative structured item judgment process (Jaeger, 1982)

- ❖ **Jaeger (1982) proposes a process that combines features of holistic judgment, examinees' performance and item judgment methods.**
- ❖ **To apply this method it is first necessary to identify appropriate population groups of examiners and then to select representative samples from each of the population groups.**
- ❖ **For example, high school teachers, tertiary institution lecturers, curriculum developers etc.**



# Setting Standards for Senior Certificate

- ❖ Each examiner is asked to complete the examination on a given subject and rate each item on a yes-no scale in answer to the question,
- ❖ “Should every regular high school graduate in the country be able to answer this item correctly?”
- ❖ The recommended passing scores for each examiner is determined by summing over the items (with a score of 1 point awarded for each “yes” response and 0 awarded for each “no” response).



# Cont.

- ❖ **The distribution of passing scores is determined for each group and the yes-no response distribution of each group for each item is tallied.**
- ❖ **Then the examiners are informed about how a sample of eleventh-grade examinees actually performed on the item (difficulty level or p-value) and how other examiners in their group had rated the item.**



# Cont.

- ❖ **Examiners are allowed to modify their original ratings in light of this information in two successive cycles.**
- ❖ **The median value of the distribution of passing scores recommended by each group is used as the standard for that group.**
- ❖ **The mean of the median values gives the standard score (cut-off) for the particular level of performance.**





# Concluding remarks

- ❖ **It is imperative to recognize that although standard setting is an important psychometric problem, it is not solely a technical one.**
- ❖ **The consequences of appropriate or inappropriate standards for individuals, for institution and perhaps for society as a whole must be considered.**



# Conclusion

- ❖ **Is there are a legitimate need for establishing a performance standard for interpretation of the examination results in question?**
- ❖ **Identify the likely threats to invalidity of the inferences that are to be made from the examination results.**



# Cont.

- ❖ **Use two or more different approaches to standard setting and multiple samples of examiners.**
- ❖ **Examine empirical evidence of how a typical sample of examinees performed on the test and to use this information in evaluating the consequences of setting a particular standard.**



# Cont.

- ❖ **The standard setting process can be done at the national level and by identifying the key stakeholders of the Grade 12 external examination.**
- ❖ **The standards so derived could then be used by provincial departments of education(schools, colleges, adult learning centres etc) for classroom instruction, by the Exam Boards in setting examination papers and by UMALUSI in issuing certificates.**



□ ***NGIYABONGA KAKHULU UKUNGILALELA  
KWENU***

