

4692

DIF analysis with multiple language groups: Theoretical possibilities and practical challenges

Paper presented at the 5th International Test Commission
Conference
Brussels, 6 - 8 July 2006

Anil Kanjee
National Education Quality Initiative

SOCIETY FOR THE IMPROVEMENT OF QUALITY



Purpose

- To identify methods/designs for conducting DF studies with multiple groups in South Africa that are:
 - most effective and practical
 - least effective and practical

Two phases

- Review of methods/designs
- Application (pilot) methods/designs

Presentation focuses on Phase 1 of project

CONCLUSION

- Recommended DIF procedure when comparing multiple groups in South Africa:

Data/Method	parametric	non-parametric
dichotomous	Logistic Regression	Mantel-Haenszel
polytomous	Polytomous Logistic Regression	Generalised Mantel-Haenszel

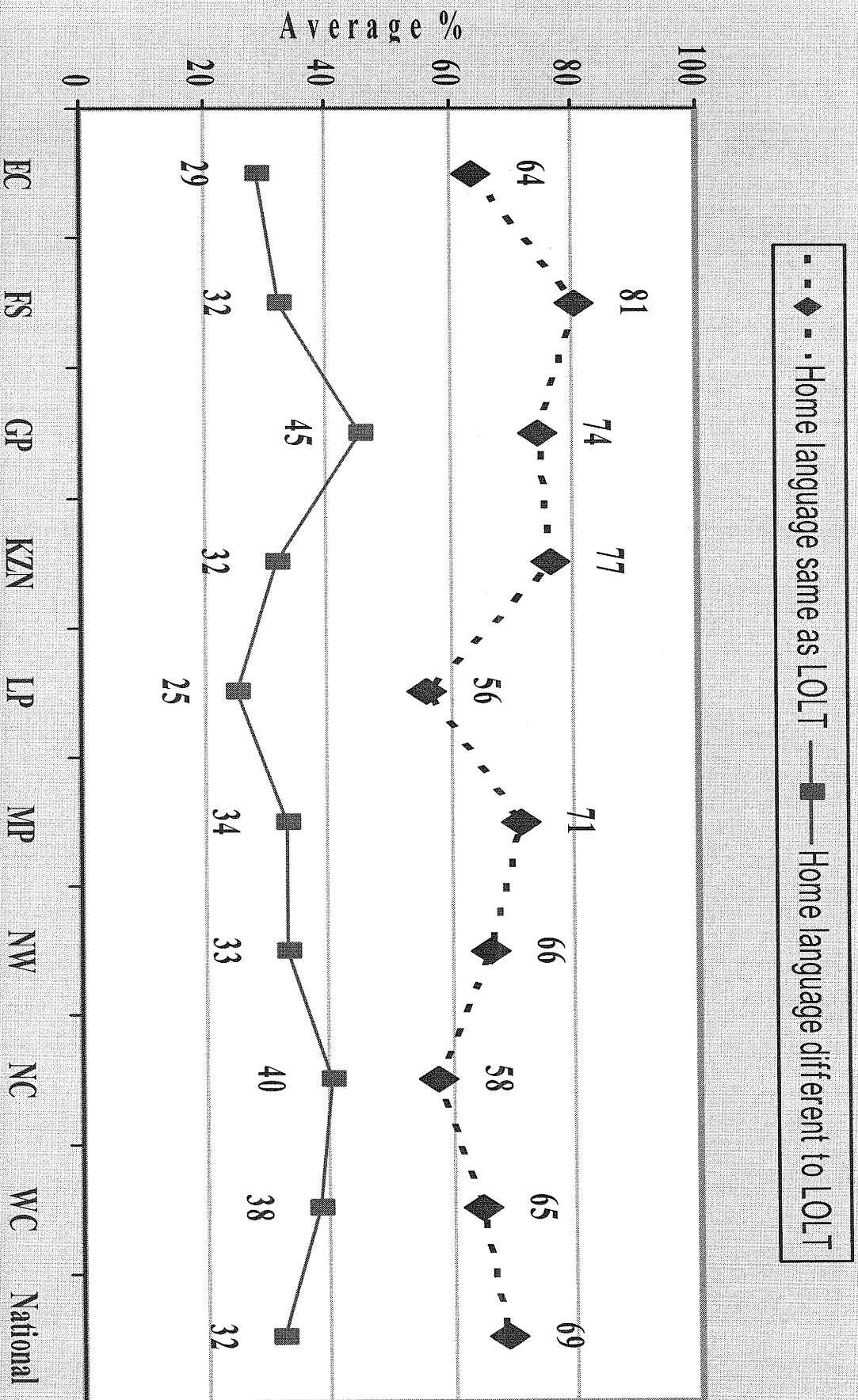
Outline of presentation

- **Background**
- **Assessment in SA**
- **Overview – DIF studies in SA**
- **Analysis Requirements in SA**
- **Ideal DIF methods/designs**
- **Index of effectiveness & practicality**
- **Issues and Challenges**

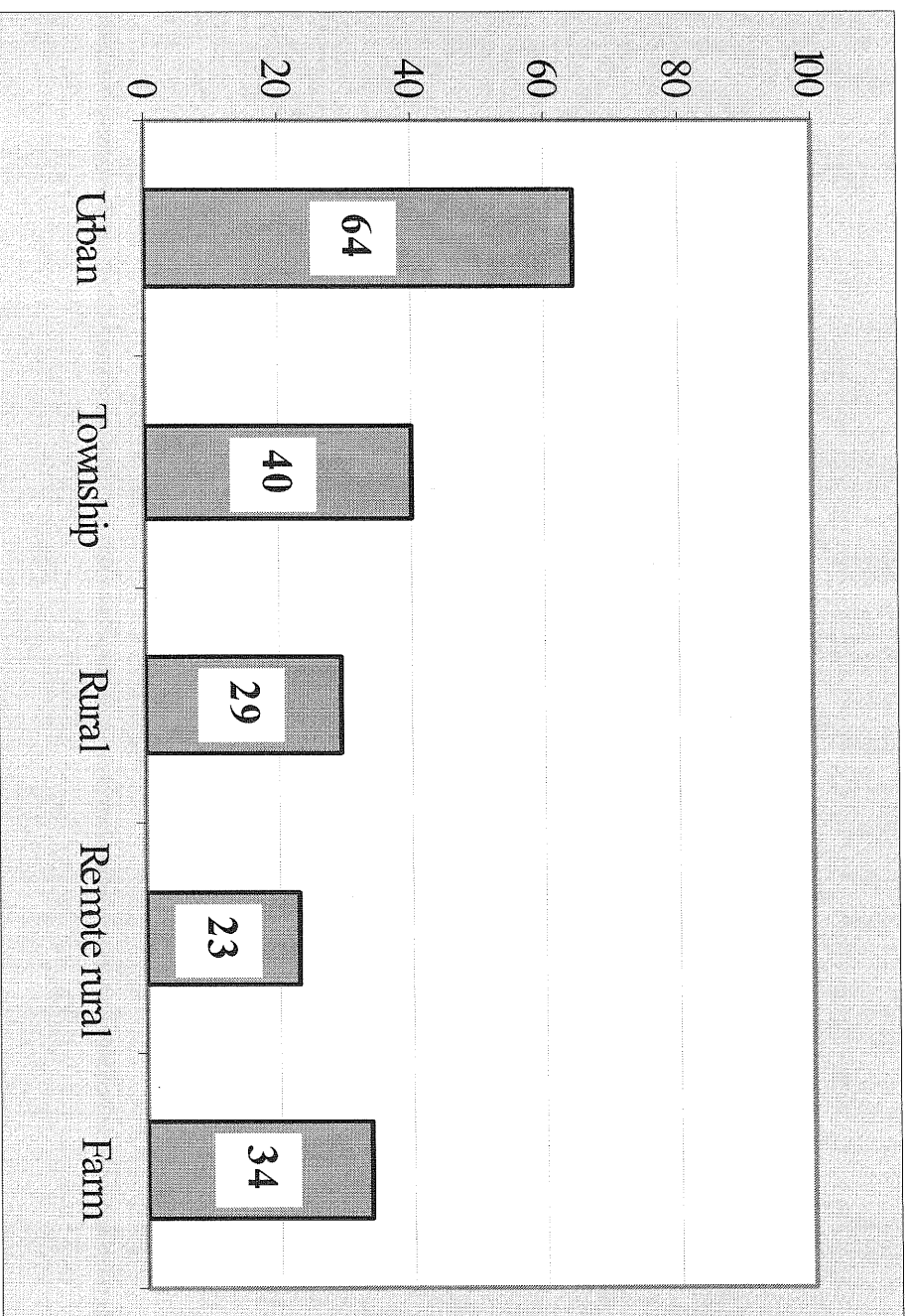
Background to problem

- Language question – perhaps the most critical question in the education sector
- 11 official languages (regionalised) – up to Grade 3
- Western Cape –introduce to Grade 6
- Instruction in second language for majority of learners ($\pm 75\%$)
- Language – significant impact on achievement
- Language – proxy for race

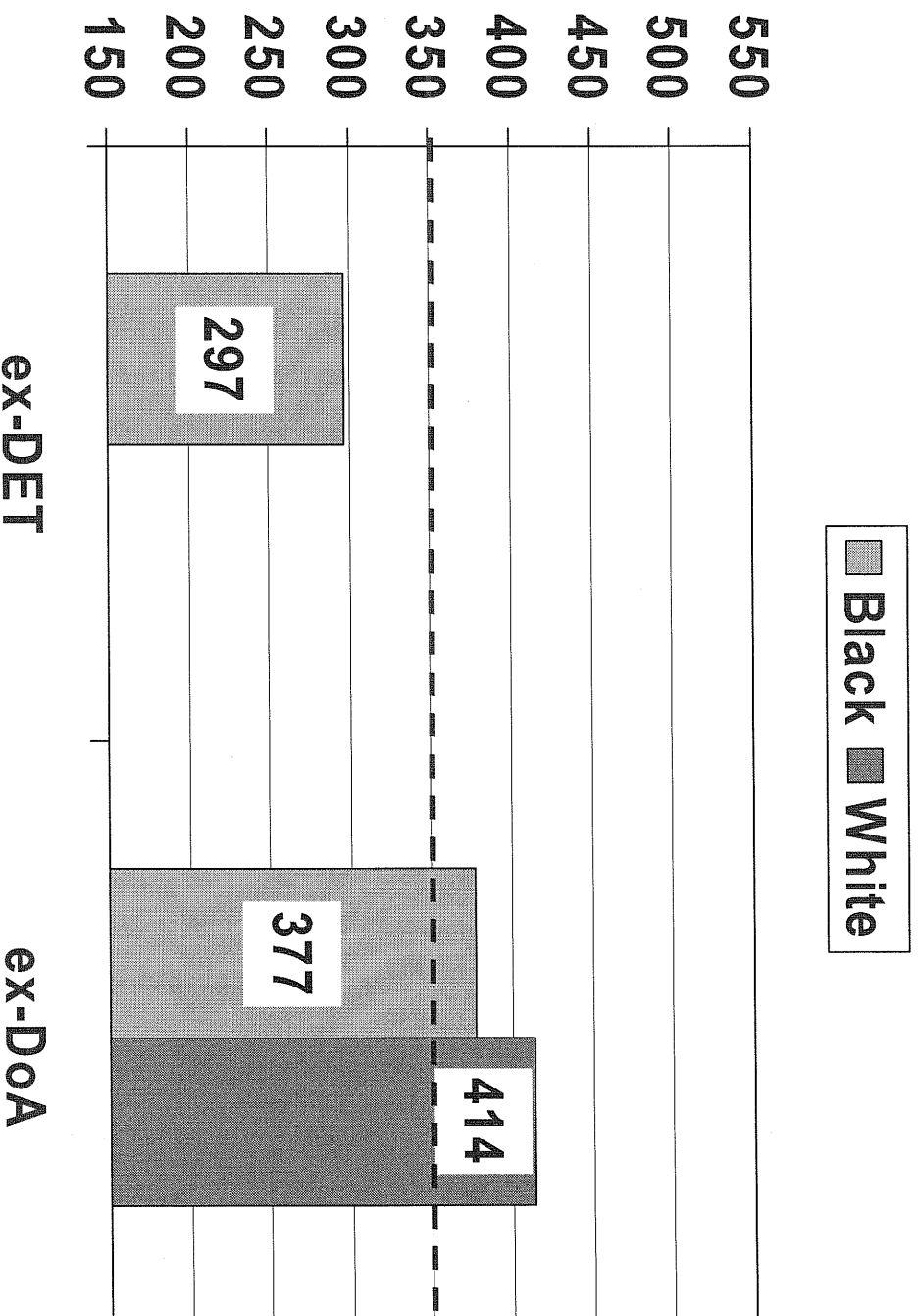
Effect of home language on Achievement: Grade 6 – National Achievement



Grade 6: LOLT by Geographical Location



Grade 9 LOLT scores by race & ex-dept.



Mean - 350; sd - 50

Assessment in SA

- Significant increase in education tests - especially large scale assessments
- Assessment community –
 - Small (high demand & pressure)
 - Limited technical expertise & experience (impact on advanced methods)
 - Limited research & publication (sharing of information)
- Nature of education tests:
 - Predominantly Free response questions
 - Mixture of Multiple-choice and Free-response
 - “Testlet” type questions – relatively common
- Limited psychological test development
 - Still primary sources of technical expertise

Context: Research/Funding

- Limited understanding of technical demands pertaining to assessment
- Extremely limited timeframes
- High demand for immediate results
 - “One button” phenomena
- Requirement of “simple” & “easy-to-understand” methods and designs
- General under budgeting of time in projects

Overview – DIF studies

- **Purpose:**
 - Address fairness/equity & validity concerns -
- **Approach:**
 - depends on data and purpose of DIF study
 - Summarized (Parametric/Non-parametric & conditioning variable Potenza & Dorans (1995))
- **Methods and techniques:**
 - Wide range, see Hills (1989), Clauser & Mazor (1998) & Penfield & Lam (2000)
 - **Primary focus – two group comparisons**
- **Current Challenges**
 - DIF + information on possible causes
 - Considerable progress has been made in the development and refinement of statistical methods for identifying items showing DIF but the development and refinement of substantive methods designed to aid with the interpretation of these items have lagged far behind **(Mark J Gierl, 2004)**

Simultaneous DIF in multiple groups

- Limited research/application
- Penfield (2001) – Generalized Mantel Haenszel
- Kim, Cohen & Park (1995) – Q_j statistic (IRT – Lord's chi square method)
- Kanjee (1995) – Pseudo IRT & LR – based on redefinition of focal group as composite group excluding reference
- **Based on dichotomous items**

DIF studies/research in SA

- Relatively few studies - thus far
 - IRT Kanjee, A. & Van Eeden, R. (1998); Kanjee & Claaseen (2002); De Beer (2004)
 - CTT Kanjee & Claassen (2000); Meiring et al (2005); Hanslow (2005)
- DIF studies not common and not standard practice (even when new tests are being developed)
- Probably others but not yet published
- To be honest, very few bias studies have been undertaken and little effort has been put into adapting mono-cultural, Westernized tests for the many cultural groups in Africa.

Prof Cheryl Foxcroft (2002) – on Choosing appropriate tests

Analysis requirements

- Sub-groups
 - 11 official languages (e.g. Grade 3 national assessment)
 - Multiple racial groups – at least 4
 - Gender
 - Geographical location – e.g. urban, informal settlement, rural, farm
- Item types –
 - free-response + MCQ
 - essay questions
 - “testlets” especially language assessments
- Designs & Methods
 - Easy to understand - funders and policy makers
 - Application: uncomplicated & quick for researchers
 - Cost effective – money & time – e.g. cost of programs
 - Easily accessible for use e.g. software

Ideal DIF method/design ?

- Testlet based items
- Simultaneous multiple group comparisons
- Different item formats – Poly & Di
- Easy application & Inexpensive
- Identify possible sources of DIF

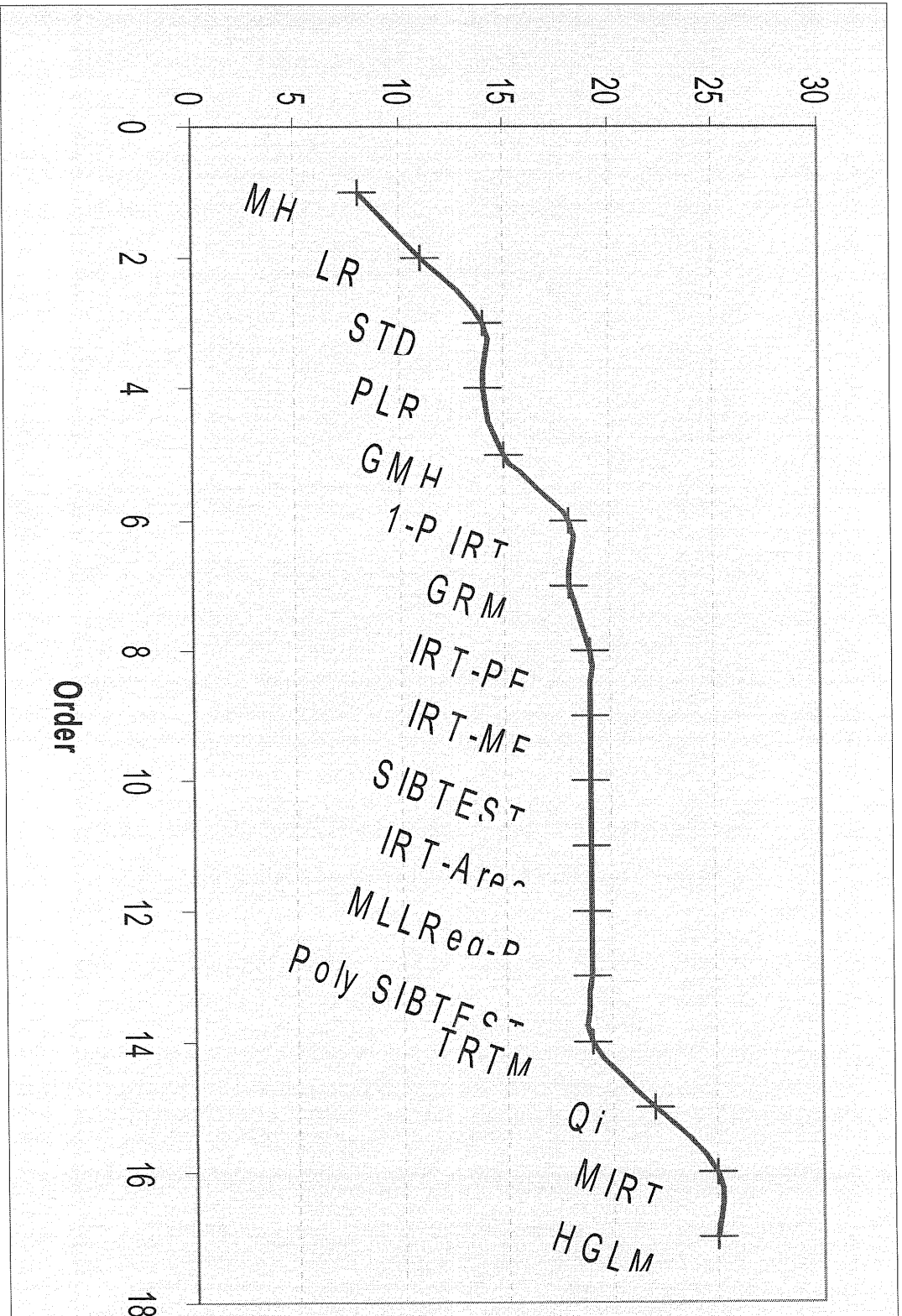
Reality – does not exist

- Which methods/designs most appropriate given specific South African context to:
 - Identification of DIF?
 - Explanation of possible causes?

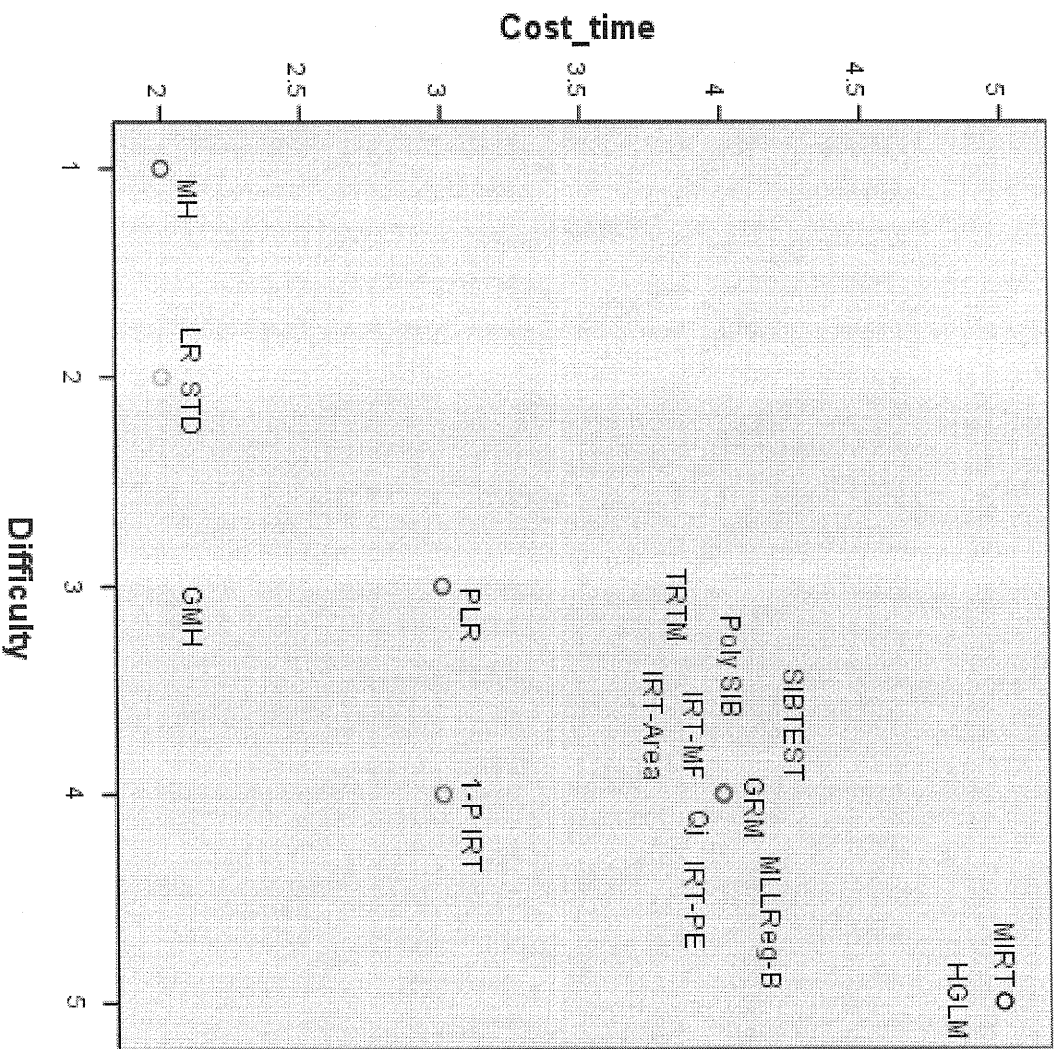
Calculation of E&P index

Procedure	Proc	Final index	Diff-culty	Cost/ time	software avail-ability	software applica-tion	use in practice
Mantel Haentzel	MH	8	1	2	2	2	1
Logistic regression	LR	11	2	2	3	2	2
Standardisation	STD	14	2	2	4	3	3
Polytomous Logistic Regression	PLR	14	3	3	2	3	3
Generalised MH	GMH	15	3	2	4	2	4
One-P model	1-P IRT	18	4	3	4	4	3
Graded Response Model	GRM	18	4	4	4	3	3
IRT parameter estimates	IRT-PE	19	4	4	4	4	3
IRT model fit	IRT-MF	19	4	4	4	4	3
Simultaneous Bias Test	SIBTEST	19	4	4	4	4	3
IRT Area between ICCs	IRT-Area	19	4	5	4	4	2
Multi-lève linear regression	MLLReg	19	4	4	2	4	5
Polytomous Simultaneous Bias Test	Poly SIBTEST	19	4	4	4	4	3
Testlet Response Theory Model	TRTM	19	3	4	3	4	5
IRT Q _j statistic	Q _j	22	4	4	5	4	5
Multi-lève IRT	MIRT	25	5	5	5	5	5
Hierarchical Generalised Linear Model	HGLM	25	5	5	5	5	5

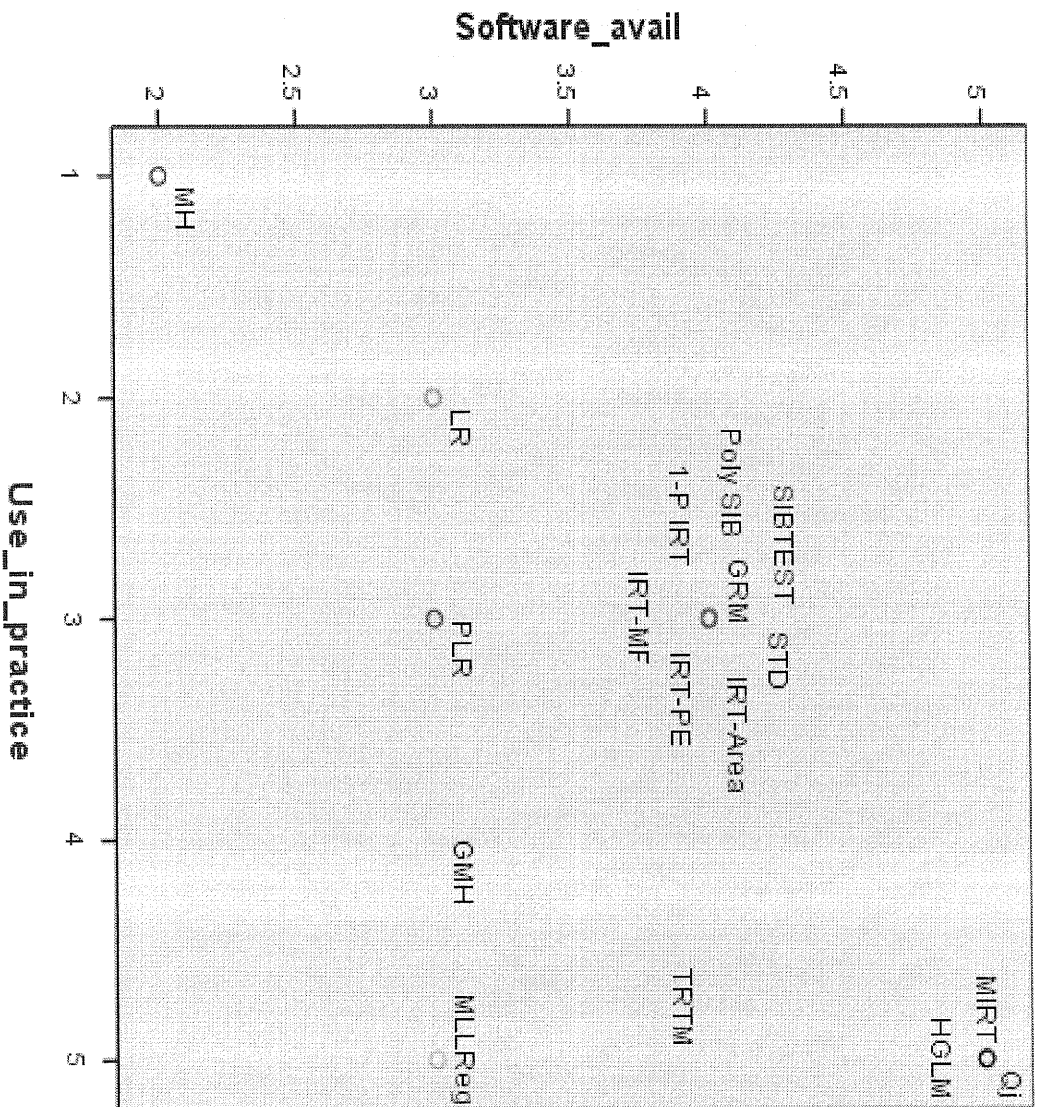
Effective & Practical Index



Cost/time vs Difficulty



Software availability by Use



Issues & Challenges in SA

- Identify possible sources of DIF
 - Instructional implications

- **Current reality**

- Functioning in an environment of severe constraints
- Until we address this AND we have a vibrant measurement culture/community,
- We have to recognize and acknowledge the context and take this into account when conducting DIF studies in general and with multiple groups in particular

CONCLUSION

- Recommended DIF procedure when comparing multiple groups in South Africa:

Data/Method	parametric	non-parametric
dichotomous	Logistic Regression	Mantel-Haenszel
polytomous	Polytomous Logistic Regression	Generalised Mantel-Haenszel

Thank you