

---

# **Good Practice in Designing Impact Studies**

**Dianne Wall and Tania Horák  
Lancaster University**

**EALTA 2007**

---

---

# What is Impact?

- Change that occurs when a high-stakes examination is introduced into an educational system
- Impact may be intended or unintended
- Impact may occur at a number of levels
  - society
  - teaching institutions
  - classrooms ('washback')
- Impact may occur at different times

---

# What is an Impact Study?

A study carried out in order to

- determine whether a new (or revised) examination has had the impact that was intended
- determine whether it has had any impact that was not intended
- feed relevant information back into the examination development process.

---

# Impact Study – Traditional View

Two phases:

- Baseline phase – to describe the situation in place before the new examination is introduced
- Final phase – to describe the situation after the examination has been embedded in the system

---

# Impact Study – Expanded View

Three phases:

- Baseline phase
- Transition phase – to track change as users of the new examination are becoming familiar with its requirements and making plans to respond to them
- Final phase

---

# EALTA Guidelines for Good Practice

Considerations for test development in national or institutional testing units or centres

## 6. Washback

- Is the test intended to initiate change(s) in the current practice?
- What is the washback effect? What studies have been conducted?

---

# Survey on Impact Studies

November 2006

## Survey of EALTA members via discussion list

- Have you been involved in the introduction of a new exam or an examination reform?
- What was the intended impact of the new/reformed exam?
- Was an impact study of any sort conducted (however small)?
- Methodology – who, when, what, how?

---

## Survey on Impact Studies

### Reasons given for not conducting an impact study

- Lack of awareness of value

‘The new system is simply assumed to be a substantial improvement ...’

- Lack of interest

‘...we can’t imagine who would have been interested in the results except perhaps ourselves.’



---

## Reasons given for not conducting impact studies

- Lack of expertise
  - ‘...because there is no expertise available and there is not a scientific language testing tradition in the region.’
- Lack of time and energy
  - ‘We knew it would be hard enough to introduce reforms; we did not have the time or energy to conduct any further tests of the kind you mentioned.’

---

## Reasons given for not conducting impact studies

- Lack of resources
- Fear of results

‘Ministries are notoriously reluctant to let outsiders see if their measures have measurable outcomes.’

‘...fear to stir the hornet’s nest – the reality might be too terrible.’

---

## Summary

Lack of

- Awareness of the value
- Interest
- Expertise
- Time and energy
- Resources

and

Fear of results

---

# Case Study: TOEFL Impact Study

internet-Based TOEFL (iBT)

Central & Eastern Europe

Qualitative study

Longitudinal (2003 to 2008)

3 stages

---

# Baseline Phase

## Pre-requisites:

- description of intended impact

- description of exam itself

- independent view of what impact was likely to occur

## We used:

- Face-to-face interviews - teachers, directors of studies and students

- Observations of teaching – TOEFL and non-TOEFL classes

---

# Transition Phase

- to track the users as they learned about the new exam
- to try to establish whether there was an 'evidential link' (Messick 1989)
- to follow up themes emerging from Baseline

We used:

- Tracking questions – every month
- Tasks
- Follow-up 'interviews'

(all data gathered via computer-mediated communication)

---

## Final Phase

Allows comparison with baseline situation.

We are using

- Face-to-face interviews – teachers, directors of studies
- Observations of teaching – TOEFL classes only.

Revised instruments – same basic idea as in Baseline, but incorporating new insights

---

## Lessons we have learned about impact studies

It is useful if the **test designers**

- state clearly what type of impact is intended.
- decide what means are to be used to encourage positive impact.
- describe the new test in detail.
- indicate clearly how it differs from any versions it may be replacing.



---

# Lessons we have learned about impact studies

The **impact researcher** needs to

- Predict what impact is likely to occur.
- Decide on timing.
  - How far in advance should the baseline study occur?
  - How long afterwards should the final stage occur?
- Design the instruments – drawing on test designers' intentions, impact predictions, and other impact studies.
- Use whatever means possible – preferably face-to-face, but also via computer.
- Disseminate the results in as many formats as possible, as widely as possible (during transition as well as at the end).

---

# Summary

- Impact studies should be an integral part of the test design process.

(How will you know what impact the exam is having without one?)

- Impact studies can be designed to fit local resources and circumstances.

---

# References

Messick, S. (1996) Validity and Washback in Language Testing. *Language Testing*, 13(3), 241-256.

Wall, D & T Horák (2006) *The Impact of Changes in the TOEFL® Examination on Teaching and Learning in Central and Eastern Europe: Phase 1*. TOEFL Monograph 34. Princeton: Educational Testing Service.

Wall, D & T Horák (2007) Using Baseline Studies in the Investigation of Test Impact. *Assessment in Education* 14/1: 99-116.

---

For more information please contact

Dianne Wall      [d.wall@lancaster.ac.uk](mailto:d.wall@lancaster.ac.uk)

Tania Horák      [t.horak@lancaster.ac.uk](mailto:t.horak@lancaster.ac.uk)