International Education Journal

Volume 2 Number 1
March 2001

Published by

Shannon Research Press
Adelaide, South Australia
ISSN 1443-1475
http://iej.cjb.net

International Education Journal

Volume 2, Number 1, March 2001

Welcome to the International Education Journal. This is a broadly based journal encompassing research and review articles. Education is interpreted in a wide manner and includes human development, learning, school education, formal and informal education, tertiary and vocational education, industry training and lifelong learning.

Online - While there will be a hard copy version of the journal it is intended that it will be accessed in almost all instances via the web.

Peer Review - All articles appearing in this journal have been subjected to a peer-review process.

Copyright - Copyright of material produced in this journal rests with individual contributors. Apart from fair use such as brief quotation for scholarly purposes, permission to use material in this journal should be obtained from the authors concerned.

The views and styles expressed in the articles in this publication are those of the individual authors and are not necessarily shared by the reviewers, the editors or the editorial advisory board.

Published by:
Shannon Research Press,
13 Endeavour Court,
North Haven, SA 5018,
Australia

Copyright © 2001

Shannon Research Press
ISSN 1443-1475

All rights reserved. Apart from individual use, no part of this publication may be reproduced in any form or by any means without prior written permission from the publisher.

Designed by Katherine Dix

Printed in Adelaide
March 2001

CONTENTS

FREE Online Subscription
Subscription is available as a free service in online format only. Either complete the subscription form at: http://iej.cjb.net or email the onlineeditor@iej.cjb.net. The reader will be notified of the latest issue and contents.

Publication Frequency
It is intended that the journal be published three times per year. However given the electronic characteristics of online publishing it is not necessary to wait until a full journal is assembled prior to publishing. Articles once accepted in appropriate form may be published without delay. It should be noted however that in electronic format pagination may vary according to the parameters of the downloading computer. Alternatively the journal article can be downloaded in PDF format, in which case the original pagination will be preserved.

Editorial Board

Editors in Chief

Dr Reaburn Reynolds
Flinders University (Education)
Bedford Park, South Australia
Phone: 08-8201 2807,
Fax: 08-8201 3184,
reaburn.reynolds@flinders.edu.au

Prof John Keeves
Flinders University (Education)
Bedford Park, South Australia
Phone: 08-8201 2392,
Fax: 08-8201 3184,
john.keeves@flinders.edu.au

Online Editor

Ms Katherine L. Dix
Flinders University (Education)
Bedford Park, South Australia
Phone: 08-8201 2105,
Fax: 08-8201 3184,
katherine.dix@flinders.edu.au

Editorial Advisory Board

Prof Jonathan Anderson
School of Education
Flinders University

Prof Bob Catley
Department of Political Studies
University of Otago

Prof Kevin Marjoribanks
Graduate School of Education
University of Adelaide

Dr Larry Saha
Department of Sociology
Australian National University

Aims of IEJ

The aim of the International Education Journal is to publish articles that possess one or more of the following characteristics or qualities.

1. The presentation of substantive findings which are of importance for policy and practice beyond the country in which the research was conducted.
2. The integration of education with academic disciplines such as anthropology, demography, economics, history, law, linguistics, philosophy, political science, psychology and sociology, or examines educational issues from the perspective of the disciplines or investigates issues at the interface between education and one or more of these disciplines.
3. The examination of educational issues from a cross-cultural or indigenous people's perspective.
4. The evaluation of educational policy or programs or the use of information technology of cross-national interest and significance.
5. The employment of advanced research methods and measurement procedures which are clearly explained.
6. The presentation of empirically or analytically based investigations of theory, models or conceptual frame works in the field of education.
7. The syntheses of research findings from comparative and cross-national studies in education.

Invitation

Contributors - Authors are invited to submit material to this journal. As a general guide articles should be not more than 5000 words. The journal may publish longer works such as theses or disserations as occassional mongraphs. In every instance material must be received in publisher ready format. Full details of publication style and other requirements are set out under the heading Author Information.

Publishers - If you wish to have books or educational software reviewed and published in the International Education Journal please contact the editors.

Reviewers - If you are interested in contributing to the journal by writing a review article (500-1000 words) please contact the authors indicating your areas of interest. We look forward to hearing from you.

Author Information

IEJ welcomes practical and research manuscripts that focus on educational issues, provide a clear purpose and depth of discussion, and are presented in a straight-forward style accessible to an international audience.

The initial submission:

- should be prepared using the downloadable Microsoft WORD97 template document (iejbase.doc), complete with layout styles, images and tables correctly positioned in the text
- images that are imported or inserted into the text from an external source as figures should be included as separate graphic files (either GIF or JPG format)
- should be forwarded electronically, either by:

  e-mail to: onlineeditor@iej.cjb.net

  or,

  on a 1.4MB 3.5 inch floppy diskette to: IEJ Online Editor, 13 Endeavour Court, North Haven, SA 5018.

The editors reserve the right to return poorly edited or improperly formatted manuscripts. Hard-copy only submissions cannot be accepted or returned.

The manuscript should:

- show the article's title, authors' names, affiliation, and email addresses (these will be included on the online page), and postal address and telephone numbers (these will be removed from the final online page)
- abstract should not exeed 150 words
- paper should not exceed 5,000 words (20,000 for occasional papers)
- include an abstract of about150 words
- include 5 keywords describing article
- be in final form ready for reviewing
- conform to the APA Reference System

General Principles for layout:

- Papers should be prepared using the template *iejbase.doc* with Microsoft WORD on Mac or PC
- Authors with WORD 98 or later and a basic understanding of web page construction should also save their paper in web format (**.html*) incorporating any hypertext features.
- Use end-notes, with one set of numbers running through the hole article.
- Use italics (not underline) for emphasis.
- Do not create additional headers and footers in the template document *iejbase.doc*, just modify appropriately.
- Do not use page numbers in text (web pages do not follow conventional page numbering systems).

If you have any questions regarding layout and presentation, please contact the IEJ Online Editor by email on: onlineeditor@iej.cjb.net

FUIIE

The publishers have arranged for the Flinders University Institute of International Education to produce and host an electronic mirror site containing all material in the Journal, accessed free of charge. There are no restrictions on downloading or printing single copies. The electronic version also provides readers with a search tool.

Partnerships 21

Paul Kilvert
Director Partnerships 21 Taskforce

This article provides a concise account of the program being introduced into South Australian schools and educational institutions during the Year 2001. This program, referred to as 'Partnerships 21', builds on developments that have taken place during the past 30 years in the schools of South Australia and that provide increased opportunity for local management of schools and educational institutions. Eight guiding principles have been advanced to ensure that the developments under the program occur in a consistent and coherent manner. The strengthening of local management is delivered through the three key components of the program: partnerships, quality improvement and resource flexibility. It is optimal for schools and institutions to join the scheme, and differences in the management and operation for members and non-members of the Partnerships 21 program are identified.

management of schools, local management, partnerships 21, resource allocation, quality in education

INTRODUCTION

As an educational reform, local school management is not new. Its various forms are already evident in the United Kingdom, Canada and the United States of America and, more recently, in the Schools of the Future in Victoria and Tomorrow's Schools in New Zealand. Although Partnerships 21 represents the South Australian version of these initiatives, it is helpful to understand that local management in South Australia also is not a new phenomenon. Under the Education Act, 1972, school councils have operated in South Australia since 1974 and already possess considerable authority, including the provision of advice to the Principal on the school's educational program and also the formal approval of the school's budget. The Freedom and Authority memorandum in 1970 from A.W. Jones, the then Director-General, assigned significant authority to the local school to develop local teaching programs derived from broad departmental curriculum frameworks. Since the mid 1980s, schools have operated consolidated bank accounts and exercised direct responsibility for procurement and financial transaction. For the last ten years, schools have exercised a degree of local selection when filling leadership and teaching vacancies. Partnerships 21, therefore, can be best understood as a strategy for moving the South Australian education and care system forward into a more cohesive and coordinated system of local management.

IMPLEMENTATION OF PARTNERSHIPS 21

The origins of Partnerships 21 arose from the work of the Ministerial Working Party on Local Management that the South Australian Minister for Education, Children Services and Training, the

Hon Malcolm Buckby, commissioned in 1998. The Minister set the following guidelines for the Working Party's investigation and report on local school management. He charged them to develop a uniquely South Australian model of local school management that possessed the following qualities:

- educational benefits must have primacy over all others;
- schools must remain responsible to the Department of Education, Training and Employment;
- teachers in schools must remain employees of the Department of Education, Training and Employment;
- the broad curriculum goals of the Department of Education, Training and Employment must be achieved;
- incentives must exist for schools to make savings through new efficiencies; and
- disadvantaged schools must be significantly better off.

On 20 April 1999 the Premier of South Australia and the Minister for Education, Training and Employment launched Partnerships 21 as the South Australian model of local management based on the general directions of the ministerial Working Party Report. The implementation timeline allocated to develop the detail of the Partnerships 21 model, from launch to inception, was eight months. From the outset it was emphasised that the model would be evolutionary and that it would be shaped by an interactive partnership between staff in state office and leaders on local sites. Seven implementation working groups in the areas of Accountability, Community Partnerships and Governance, Professional Development and Training, IT System Support, Global Budgets, and Asset and Financial Risk Management, Human Resource issues and Country issues developed the detail of Partnerships 21. Their work was overseen by a Steering Committee that the Chief Executive chaired. Over 200 people, representing a wide cross-section of the South Australian community, worked on these groups.

PARTNERSHIPS 21 PRINCIPLES

A set of eight guiding principles were approved by the Steering Committee to ensure that the developmental work for Partnerships 21 occurred in a consistent and coherent manner. These eight principles continue to apply to its operation and development. They are:

1. Building on success

Partnerships 21 will:

- Build upon past strengths and best practices when implementing new structures and processes for the local management of sites and services.
- Strengthen the connections and strategic alliances between sites and services in achieving high quality learning outcomes for students and children and for optimising the management of their resources to do so.

2. Improved learning outcomes

Partnerships 21 will:

- Strengthen a culture of achievement in sites and services in which leaders and teachers measure, and with parents, reflect on and target action to improve student learning outcomes.
- Optimise the efficient use of resources to achieve high quality learning outcomes for students.

3. Partnerships

Partnerships 21 will:

- Strengthen participation by students, parents, community members and staff of all cultural backgrounds in an open democratic process of decision making at local sites and services.
- Promote the development of ethical partnerships with clusters and networks of sites, services, businesses and other community and departmental organisations.
- Promote voluntary strategic alliances between sites, services, businesses and the broader community.

4. Fairness

Partnerships 21 will

- Reallocate resources differentially to assist children and students in urban, rural and, in particular, isolated sites to achieve the standards and benchmarks of the South Australian Curriculum Standards and Accountability Framework.
- Reallocate resources differentially to assist all children and students in urban, rural and, in particular, isolated sites gain access to pathways that best assist their development and success.

5. Maximising the local

Partnerships 21 will:

- Maximise decision making at the local level unless there are compelling reasons for state office to make resourcing or other decisions.
- Remove any unnecessary restrictions on the flexibility of local sites and services to optimise their resources to ensure high quality learning for their students and children.
- Strengthen the state office's service role to sites and services.

6. Accountability

Partnerships 21 will:

- Strengthen the ability of sites and services to provide their local communities with information that explains their performance in providing high quality care, education and training against agreed benchmarks.
- Strengthen local sites and services' ethical and professional use of performance information and benchmarks as part of their process of self-evaluation and continuous improvement in student learning outcomes.
- Strengthen the ability of sites and services to provide state office and the people of South Australia, through the Parliament, with information about their performance in providing high quality care, education and training.

7. Efficiency

Partnerships 21 will:

- Optimise the efficient use of resources at the local and central levels of operation.
- Strengthen decision-making practices and management structures for the efficient use of resources both in and between local sites and services.

8. Transparency

Partnerships 21 will:

- Ensure systems and processes that service local management at state office are open, defensible and ethical.
- Ensure sites and services provide processes and information that are accessible to all groups within their communities and that are open, defensible and ethical.

KEY COMPONENTS OF PARTNERSHIPS 21

From the outset it has been emphasised that Partnerships 21 is committed to strengthening the state system of education and care. The following requirements are made clear in the model.

- Partnerships 21 schools and preschools continue to operate within the legislative requirements and departmental policies that apply to all sites.
- Partnerships 21 schools and preschools continue to teach programs consistent with the department's broad curriculum goals as defined by the South Australian Curriculum Standards and Accountability Framework.
- Children and students continue to have right of access to their local neighbourhood school or preschool.
- The department remains the employing authority.
- Increased staffing flexibility occurs within current industrial agreements.

Within these requirements, the Partnerships 21 model provides a coherent framework for strengthening local management of schools and preschools through the key components of partnerships, quality improvement and resource flexibility.

Partnerships*:* The Partnerships 21 project builds on South Australia's long history of involving of parents and community members in schools and preschools through school councils and preschool management committees. It extends this practice by redefining concepts of governance, management and leadership in the context of local management in school and preschool communities and the system as a whole. The focus for this increased emphasis on partnerships with the local community is the strengthened role of the school's governing council. Under Partnerships 21 its role is extended from giving advice to the Principal to one where it exercises authority for strategic planning, policy determination, monitoring and accountability.

Quality improvement*:* The implementation of Partnerships 21 has been tied to the department's introduction of a new Framework for Quality Improvement and Accountability. The new framework is based on a three-year strategic plan, known as the Partnerships Plan, which is an extension of its predecessor, the annual Statement of Purpose. The Partnerships Plan is a companion document to the Services Agreement, which identifies the mutual obligations that Partnerships 21 schools and state office have of each other. The Partnerships Plan and Services Agreement, are signed by the chairperson of the governing council, the Principal and the Chief Executive. The other key features of the Framework for Quality Improvement and Accountability are annual operation planning, monitoring and accountability.

Resource flexibility*:* Under Partnerships 21, schools are allocated a global budget and the authority to determine how it will be administered. The global budget allocates resources to a school, based on a new allocative mechanism using the following components: a base allocation for each site; a base allocation for each student according to their year level; and additional allocation for students with special learning needs, for example, Aboriginal students, students with disabilities, English as second language students, students from low socio-economic backgrounds; and funding for special programs that are unsuitable for a per capita allocation.

Through the integration of these key components, the school or preschool is given the best opportunity to improve the learning outcomes for each student. The panels that follow summarise the relationships between the Governing Council of an Institution and the Principal or Director of the Institution and the differences in management and operation between the Partnerships 21 and Non-Partnerships21 schools and institutions.

PARTNERSHIPS OVERVIEW

Panel 1. Governance

Governing Council	Principal / Director
Is accountable to the Minister for developing, negotiating and meeting the objectives and targets of the Partnerships Plan and Services. Agreement by: · strategic planning and the allocation of resources, · monitoring key indicators and levels of client satisfaction, · reporting to department and community. Is responsible for local policy development within broad DETE frameworks (eg curriculum and program initiatives) Participates in the appointment of key leadership positions. Has employer responsibility for staff employed by the governing council.	Is accountable to the Chief Executive through the designated Executive Director for the educational leadership and management of the site (DETE sites only) and the development of the Services. Agreement resulting in: · quality teaching programs, · quality learning outcomes for all students, · a safe learning and working environment, · effective operational and day-to-day management, · supervision of all staff on site. Is accountable to the governing council for the: · implementation of the Partnerships Plan and Services Agreement, · implementation of local policy, · provision of accurate and timely information and advice, · supervision and development of staff employed by the governing council (as delegated).

Panel 2. Accountability

Governing Council	Principal / Director
Is accountable to the community and the Minister for meeting the objectives and targets of the Partnerships Plan and Services Agreement. Is responsible for consultation with the community and ensuring that decisions take into account the range of community views, with particular attention to the views and needs of disadvantaged and minority groups. Is accountable to the community for ensuring that the educational needs of students are addressed.	Is accountable to the Chief Executive for the education outcomes of students. Is accountable to the Chief Executive for the educational leadership and performance management of staff by: · working with staff to develop and deliver quality educational programs which meet students' needs, · reporting to the school community on curriculum developments, · ensuring the provision of quality training and development for staff. Is responsible for monitoring and reporting on student learning outcomes to the Governing Council and Chief Executive.

Panel 3. Monitoring and Evaluation

Governing Council	Principal / Director
Monitors progress of the Operational and Partnerships Plan. Monitors the budget and variations to the budget. Monitors the current and future financial position. Oversees and participates in annual internal reviews. Participates in external reviews. Meets information and accountability requirements.	Reports to Governing Council on key participation, learning, training and care outcomes. Provides timely and accurate information and regular reports on issues relevant to the responsibilities of the Governing Council. Is responsible for performance management of staff including staff development and the allocation of duties. Manages annual internal reviews.

Panel 4. Setting the Strategic Direction

Governing Council	Principal / Director
Is accountable for setting of the broad direction and vision of the school/preschool and for the development, monitoring and review of the Partnerships Plan. Ensures the inclusion of DETE priorities within a local context. Approves the Annual Operational Plan and endorses the budget and variations thereto. Enhances and supports the program of the school/preschool through, for example: · alliances and contracts with business and other agencies and services, · alliances and contracts with other DETE sites and services, · fundraising and sponsorship within nationally agreed guidelines, · setting materials and services charges.	As educational leader, provides advice and information to the Governing Council to assist in the development of policy. Is responsible for the development of operational policy and action plans to achieve the objectives and targets of the Partnerships Plan. Is responsible for financial management and expenditure consistent with the Annual Operational Plan and approved budget.

Panel 5. Setting Local Policy

Governing Council	Principal / Director
Establishes local policy within broad DETE framework, eg: · codes of conduct for Governing Council members, · client and Governing Council employed staff grievance procedures, · financial policy (investment etc), · curriculum policy (eg emphases, priorities), · human resource policy, · facilities development and maintenance policy, · school dress code, · school discipline policy and student/children behaviour management code. Allocates resources to support local policy implementation.	As educational leader, provides advice and information to the Governing Council to assist in the development of local policy. Is responsible for financial, physical and human resource management. Is responsible for the implementation of: · school dress code, · school discipline policy and student behaviour code.

Panel 6. Quality Improvement Overview

Feature	Partnerships 21 site	Non Partnerships 21 site
Quality Improvement and Accountability Framework	A cohesive, integrated quality improvement and accountability framework expressed through a three year strategic plan, a services agreement between the site and state office, an annual operational plan and an annual report.	A cohesive, integrated quality improvement and accountability framework expressed through an annual statement of purpose and an annual report.
Partnerships Plan	A three year strategic plan incorporating staffing and facilities planning.	An annual Statement of Purpose.
Services Agreement	A formal commitment of resources for one year and a projected allocation for a further two years through a services agreement between the local site and state office.	Formal commitment of resources for no more than one year.
Annual Operational Plan	Annual operational plan integrating resource planning with strategic priorities.	School improvement / action plans for specific purposes as required.
Annual Report	Annual Report	Annual Report
Internal Monitoring	Individual sites design processes to monitor progress against strategic objectives outlined in the Annual Operational Plan and Partnerships Plan.	Individual sites design processes to monitor progress against annual priorities.
External Monitoring	Basic Skills Test. SACE results. Statements and Profiles state wide collection of data. Site review every three years. Site review for specific purposes on request.	Basic Skills Test. SACE results. Statements and Profiles state wide collection of data. Site review every three years. Site review for specific purposes on request.
Curriculum Standards and Accountability Framework	Progress of students will be identified, monitored and built upon.	Progress of students will be identified, monitored and built upon.

Panel 7. Financial Management

Feature	Partnerships 21 site	Non Partnerships 21 site
Funding Mechanism and Accountability	A global budget based on a transparent, predictable, output-based formula. Accountability integrated into the Quality Improvement and Accountability Framework.	Many different formulas related to inputs with separated and unrelated accountability measures.
Resource Allocations	P21 sites contribute to developing a new index of educational disadvantage to target better funding to those students with the greatest need. An amount to cover gap between School Card payments and at least the maximum legally enforceable payment. Unacquitted BTS grants paid as a lump sum. Increased allocations for breakdown maintenance and utilities based on per capita formula will be built into the Global Budget. Savings retained by the site.	School Card payments remain the same. Sites will have bills paid as they arise and not be able to retain savings.
Financial Management	A one year firm budget with a 2 year projected budget. An overdraft facility, equivalent to $240 per student, will be available to Partnerships 21 sites supported by a three year cash flow plan Savings and over expenditure will be carried forward. Monthly and annual financial reports to assist planning and monitoring. Access and training to use cash flow plans and other tools that complement EDSAS finance module. Shared service centre development to share or remove workload associated with increased financial management.	Cash grants for each financial year received at different times, allocated by a range of formula and accounted for by different mechanisms.
Asset Management	Priority for the development of Asset Management Plans. Professional development related to asset management. Accountability for BTS grants built into the Partnerships 21 Quality and Accountability Framework.	No change to present procedures and policies. Sites will use the present acquittal processes.

Panel 8. Human Resource Management

Feature	Partnerships 21 site	Non Partnerships 21 site
Industrial Agreements	Existing agreements maintained.	Existing agreements maintained.
Staffing Entitlements	Calculated on basis of staffing documents, and sites given integrated global entitlement for all sites for one year with indicative rolling triennium.	Calculated on basis of staffing documents. Schools receive entitlements for one year. Preschools, each term.
Long Service Leave	Principal / Director approval of any length of time.	State office approval for periods of one term or more.
School Choice Procedures	Includes all vacancies created by completion of teacher tenure, and at least 50 per cent of any other placements.	Vacancies created by the completion of teacher tenure, plus one.
	Increased opportunity to advertise for permanent and limited tenure staff.	Limited opportunity to advertise.
Advertising Leadership Positions	Increasing ability to determine leadership structure and mix of positions. Can advertise at least 70 per cent of leadership positions that become available.	Leadership positions released for advertisement or placement following approval.
Local Choice for limited term SSO appointments	Sites can select SSOs/ECWs for positions that are of a minimum of two terms, up to 41 weeks.	Continue to be centrally managed short term appointments.
Ancillary staff classifications	Access to a greater range of SSO classifications from SSO 5 and beyond.	Maximum SSO 4.
Appointment of short-term staff	Principals can appoint staff to short term positions.	Managed by state office.

Learning through the physical environment in the workplace

K.M. (Paddy) O'Toole
Flinders University *paddy.otoole@flinders.edu.au*

The physical surroundings are often overlooked in discussions on learning in the workplace. The physical environment, however, may hold significant messages for organizational members in relation to what they need to know about the culture, structure and roles and routines of the organization. This paper discusses how differences in the physical environment of two departments in the same organization influenced the way that people worked and learned.

organizational culture, physical environment, organizational learning, organizational structure, workplace learning

INTRODUCTION

The amount of investigation of learning in the workplace has increased dramatically in recent years (Argyris, 1992/1999; Cook & Yanow, 1996; Czarniawska, 1997; de Geus, 1988; Huber, 1996; Larson-Knight, 2000). Globalization and competitive pressures have forced organizations in many sectors to attempt to increase learning at individual, team and organizational levels in order to ensure that the organization survives and prospers. It should be acknowledged, however, that besides learning about the work processes and competitive environment of the organization, organizational members need to know about the culture and structure of the organization to operate effectively. This paper contends that organizational members draw messages from their environment that supports their learning.

The nature of organizations is such that interactions take place not only between members, but also between the members and their physical surroundings, such as buildings, their layout and their objects (Assmann, 1995; Domingues, 1997; O'Toole, 2000a). Past activities and events are embedded in the patterns of behaviour and the artefacts, and have symbolic meaning for the members (Assmann, 1995; Domingues, 1997; Doxtater, 1990; Gagliardi, 1990; Gagliardi, 1996; Hatch, 1990; Larsen & Schultz, 1990; O'Toole, 2000a; O'Toole, 2000b; Rosen, Orlikowski, & Schmahmann, 1990; Witkin, 1990).

The purpose of this paper is to discuss the relationship between the physical environment of the organization and learning in the workplace. It is argued that the physical surroundings of an organization such as the building, layout, machinery, equipment, and uniforms may have a significant impact on the way people work and therefore the way they learn. It is important to note that learning need not be intentional. Learning is defined as taking place where potential behaviours change through acquiring, distributing or interpreting information (Huber, 1996). In this paper, learning is deemed to have taken place where people learn their tasks, role, the structure and culture of the organization or new ways to operate.

After the literature is reviewed, a case study is discussed where research on two different departments in a small organization took place. The particular focus in this paper is the organization's physical layout, which is investigated in the cultural and structural context of the organization.

THEORETICAL FRAMEWORK

It is clear that people create the physical environment around them. People design and build buildings, create and implement floor plans, choose and obtain the objects with which they surround themselves. At the same time, the physical environments thus chosen also influence the individuals who dwell and work within the same physical surroundings. The environment that surrounds people is an element of social structure, and, as Giddens explained, becomes part of the structuration process, influencing the humans associated with it, while human agency influences the organization and substance of the artefacts within it (Giddens, 1984; Gieryn, 2000; Rosen et al., 1990). This influence relates to meaning, values and beliefs that form part of the "place-identity" that a site or object may hold for an individual.

Proshansky, Fabian and Kaminoff (1995) saw "place-identity" as a sub-structure of an individual's self-identity. Schemas relating to an individual's past experience of how places have satisfied needs and desires form as the individual matures. Out of good and bad experiences emerge values and beliefs about the physical world and its meaning. Part of these meanings relates to role associations. An individual learns through childhood the appropriate behaviours relating to roles in school, home and neighbourhood. Environmental understandings and competence result from an individual's adaptation to each setting (Proshansky et al., 1995). As the individual matures and commences work, this adaptation and adoption of roles according to place continues.

Because place constitutes only background, perhaps it will only be when the physical setting does not meet the needs of the individual that the individual may be become acutely aware of discrepancies (Proshansky et al., 1995). Proshansky et al. gave the example of a new Associate Professor who has to share his office with many other young faculty members.

> Missing are the books, research files, private conversation area … associated with the role of academic faculty researcher …. the meanings of space and place are not being met by the actual physical setting in which he is expected to play this role …. we can expect the 'Assistant Professor' to be very much aware of his [inadequate] … physical setting, particularly when his students come to see him; and second, at some level of awareness he may self-consciously be uncertain of his status as an Associate Professor especially in regard to his colleagues who have the rank but more adequate academic offices.
>
> (Proshansky et al., 1995, p. 93)

The organization of physical space may contain potent messages for members of an organization according to the meanings attributed to them. Gargliardi (1990) described how, in the U.S. Navy, there was a clear allocation of rooms for formal lectures on the one hand, and discussion and criticism on the other. This division of space was occasioned by the need of the Navy to enforce discipline and authority while promoting discussion and critical thinking in new officers. The division of space clearly indicated the behaviour appropriate to each norm, and the recruits learned the appropriate use for each space.

Schein (1985, pp. 240 - 241) gave a very clear account of how two organizations' physical environment, in the form of their floor plan, encoded information about the culture:

> Action, with its assumptions about truth through conflict and the importance of open communications, has chosen an open office layout with partitions only high enough to permit a sense of privacy when one is sitting down. Private offices are given only to a few people who need them, and then typically have glass door so that one can always see who is in....Managers at Multi spend much more time thinking things out alone, having individual conferences with others who are centrally involved, and protecting the privacy of individuals so that they can get the work done. What the visitor encounters in this organization is a lobby manned by a guard and closed doors on all sides.

Technological objects are also manifestations of culture, both of the organization and of the broader social system. Mumford (cited in Kingery, 1993, p. 205) suggested that “the machine cannot be divorced from its larger social pattern; for it is this pattern that gives it meaning and purpose”. Kingery (1993, p. 207) argues:

> In a factory, for example, there is a system of discipline, of rules, of politics in the traditional sense. The forms of machines help enforce these rules: they suggest the easiest possibilities to those who use them. They mediate between the people who make the rules and the people who have to follow them.

The physical layout of workplaces can also affect the behaviour of organizational members (Oldham & Rotchford, 1983; Strati, 1990) and show the structure of an organization (Giddens, 1984; Rosen et al., 1990). The physical separation of offices insulates each member and gives a measure of autonomy to those within them, and also serves as a powerful marker of hierarchy (Fischer, 1997; Giddens, 1984; Rosen et al., 1990).

Organizations may choose to adapt their physical layout to promote learning that gives a competitive edge. Brenner and Connell (1994) conducted research on the privacy and collaboration needs of knowledge workers. They found that the configuration of floor plans was one factor that actively promoted a learning environment. Pedler, Burgoyne and Boydell (1991/1997) suggested that the environment of an organization may help support the learning climate. They gave the following examples of "architectural practice with the possible organizational interpretations" in Figure 1.

Design Features	**Function**
removal of dividing walls partial removal of floors	decentralize functions, remove central services
outside staircases; putting service pipes, etc. outside central courtyards, wells, atriums	encourage outside trading re-train people, encourage radical job changes
add balconies re-cycling old bricks, etc.	celebrate differences, encourage expression
use historical objects as sculpture lots of inside greenery	blur home/work/community boundaries
put skylights in the roof put in bigger windows preserve historical objects	full disclosure; open up top management processes for inspection and comment
demolish departmental boundaries	encourage secondments outside

Figure 1. Architectural features with the possible organizational function, as adapted from Pedler et al. (1991/1997, p. 127)

It should be noted, however, that Pedlar et al. make no claim to prior research confirming their conclusions, and the emphasis in their arguments should be on the word "possible". Von Krogh, Ichijo and Nonaka (2000) observed how architecture and disposition of floor space appeared to promote an organizational learning culture in a variety of firms. An example of this occurs in Phonak, a Swiss company making advanced hearing systems. Phonak House is described as "a visual representation of the company's boundary-breaking culture" (p. 37), with open areas, no separate floors for directors, and a minimum of dividing walls.

Field (1995) also saw physical layout and artefacts as factors to promote or impede organizational learning. He cited bundy clocks as a mechanism to “communicate mistrust and lack of respect” (p.

58) and saw manifestations of environment such as uniforms, partitions and differentiation in office size and associated amenities as promoting barriers to communication and working collaboratively.

Hence, physical layout, artefacts and place-identity may contain role associations, symbolic meanings, and messages about structure and power in the organization. It is also suggested that physical layout and artefacts may themselves contribute or hinder effective learning in the organization that contributes to the organization's achievement of strategic goals.

THE CONTEXT: DOVE FUNERAL SERVICES

Dove Funeral Services (Dover Funeral Services) was a family business of funeral directors that employed approximately 40 people. At the time of the study the business had a Head Office and several branches in the metropolitan area, although this paper concentrates only on the Head Office. The organization was managed by a group of four managers, lead by the General Manager.

The organization was basically structured along the following functional lines:

Administration - included wages, word processing, prepaid funerals and accounts;

Coordination - were responsible for the allocation and rostering of staff and equipment to transfers of bodies, funeral arrangements, viewings, services and interments;

Funerals - the tasks of arranging funerals, viewings and conducting funerals (including interments);

Mortuary - where the bodies were prepared for burial.

The part of the study dealt with in this paper is mainly concerned with the funeral and coordination staff.

Methods

The research methods employed were mainly qualitative, with an ethnographic focus that sought to "construct descriptions of total phenomena within their various contexts and to generate from these descriptions of total phenomena the complex interrelationships of causes and consequences that affect human behavior toward, and belief about phenomena" (Goetz & LeCompte, 1984, p. 3).

The case study was based on visits to the organization’s Head Office and one of the branches. The data were collected over a period of approximately of two months. Ten semi-structured interviews were conducted with a selection of staff and managers. The staff interviewed included people from the coordination office, the administration area and some funeral staff. The remaining staff were requested to complete a semi-structured questionnaire. Observation took place within the Head Office, and documents prepared within the organization were collected for later analysis. The different methods of data collection were employed to ensure a method of triangulation to test the validity of the data. Also, staff who were interviewed were asked to check transcripts of their interviews for accuracy and meaning and a report was sent to the General Manager of the organization for his comments and feedback.

Analysis

While some preliminary analysis was conducted during the data collection, the main analysis took place in the months after data collection was concluded. Quantitative questions were converted to tables and charts using spreadsheets in Microsoft Excel 98. The narrative portions of questionnaires were converted to NUD*ist 4 documents, as were the interviews and the observation field notes. Documents that had been collected from the organization were also noted on NUD*ist. Themes that had emerged from the research questions and the data collection were created on NUD*ist, then portions of the interviews, questionnaires, field notes and documents were assigned to each theme. As the data were analysed further themes emerged; some themes were discarded while some old themes were subsumed into these new themes.

THE FINDINGS OF THE STUDY

The findings of the study were classified at the organizational level that were relevant across the organization as a whole, and at departmental level, where the situations for the coordination staff and funeral staff were compared.

Findings: The Organization

One of the major undertakings of the study was to discover the collective understandings that underpinned the site's culture, and it is necessary to examine the culture before an attempt can be made to understand the meanings of the organization's artefacts and spaces.

A value generally held throughout the organization was that of customer service. Customer service was continually mentioned in interviews, observations and recorded in the questionnaires. There seemed little doubt that it was a core value for the company, and that it formed part of their cultural identity. Some typical comments relating to the organization's primary focus were:

> Provide a good service to the people who need it
>
> Serve our clients to the best of our ability and be the best at what we do.
>
> …. Being professional, very professional and I guess being there for the family, to help them through the arrangements, support the family in any way we can.

Many of the staff also saw the company as old-fashioned and formal in the way it conducted business:

> They are pretty set and traditional, I feel
>
> …. That’s everyday in our, our attire and the way we conduct things, the way our business is presented, so yes, its formal but it’s also formal from a managerial point of view in that the staff are not really involved in any of the managerial decisions or discussions.

This formality did not extend to the interactions between managers and staff. While the authority of the managers was not questioned, and staff were not involved in managerial decision-making, the relations between staff and managers seemed extremely cordial. All managers were called by their first names and the staff often did not use (or know) the managers' formal titles.

The Physical Environment

The Head Office was clearly separated into two areas. These areas could be respectively described as "client" and "functional".

The client areas were furnished with muted colours and upholstered furniture in a traditional style. Tissues were in evidence, contained in ornamental tissue boxes. One of the client areas was a cottage that contained four meeting rooms as well as the office of the Bereavement Education manager. At the back was the casket viewing room. This cottage was sometimes used for meetings and making arrangements, when the client wished to come to the premises. The impression gained by the researcher was of comfortable, yet formal, areas where sensitive discussions could take place.

A larger building next to the cottage also contained a client area, but was the main location of the operations. A plan is shown in Figure 2.

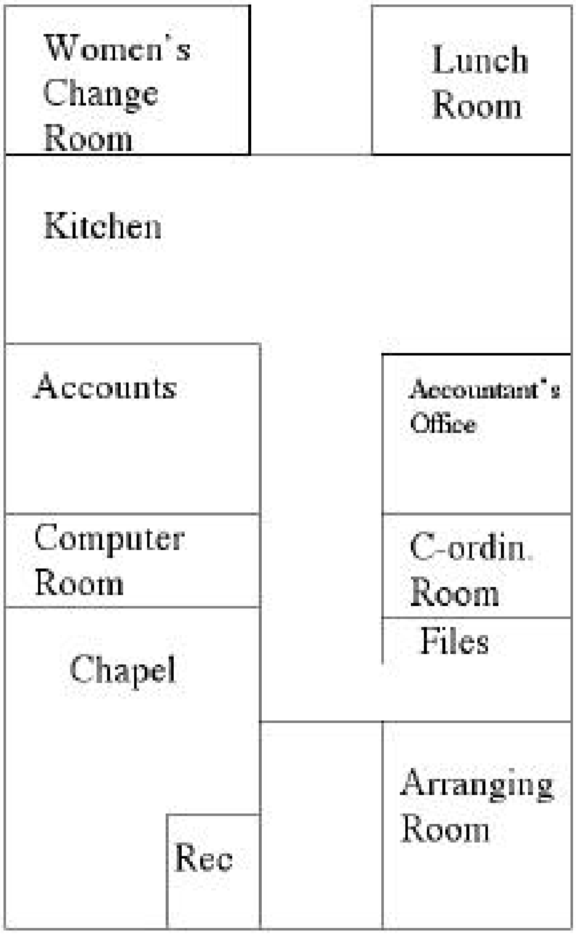

Figure 2. Floorplan of main building at Dove Funeral Services

The client area in this building was furnished in timber and decorated in dusky pink. Piped music was played. The configuration of rooms was such that funeral services could be held.

The functional area was less well maintained, although it was air-conditioned. This area was comprised of the coordination office, the accounts office, an office for the Administration Manager, the computer room, a kitchen, a lunch room, the women's change room and areas for files, photocopier and storage. The bathroom contained a bath/shower as well as the usual amenities, with a lock on the main bathroom door so that the staff were able to change from their car washing clothes to the required uniform. The funeral staff did not have desks, and meetings between the funeral staff were seen to take place in the lunch-room. The lunch-room contained newspapers and a vending machine, as well as table and chairs.

The four managers of Dover Funeral Services had their own offices, while other staff either shared work space or, as in the case of the funeral staff, had no work space allocated. The staff generally had unrestricted access to the organization's buildings, however the main staff meeting areas were the lunch room and kitchen. These two rooms were communal space where staff could hold meetings, have social chats and find out what was going on in the organization. The managers would also frequent these areas, where staff could approach them on various matters.

A Tale of Two Departments

The major difference noted between the coordination area and the funeral staff was the degree of the individualization of the work. In the funeral area, the level of discernible similarities in task performance was low, particularly in the more experienced staff. A high level of discretion was allowed by management, probably due to the complex nature of the work. The comment below illustrates the degree of individuality in task performance:

> … you could go out on a funeral with one conductor here in the company and then like it suits his style, and then another conductor does a funeral for a person, same age, same chapel, everything the same situation, and you'd swear they were two different companies.

There was some evidence that this high level of individuality made the job difficult to learn for new staff. One staff member commented:

> Funeral conductors all have their own way of conducting [therefore it is] difficult for new staff - no consistency and so therefore newer staff can be in the bad books and not know why.

When learning the job, funeral staff appeared to gather information from the more experienced staff members and then formulate their own ways of doing things. In the coordination office, on the other hand, there was a higher level of standardization, as shown by this comment:

> … there are many things we do, that we have to do because that's the way [Dover Funeral Services] does them, mostly in funeral arrangements, even to the point where, when we fold forms we fold them to the outside, and they are all folded a certain way.

The coordination office staff tended to work in pairs, in several shifts. Where possible, a more experienced person would be teamed with a newer staff member, and each staff member was seen to assist the other wherever possible. One staff member was observed handing out notes from an old file that she felt would be of assistance to the other people in the coordination office. Files were centralized so that access was available to all the coordination staff members.

The task structure differed between the coordination office and the funeral staff. While three funeral staff would be allocated to work on one funeral, it was observed that each person had a defined role, and the person nominated as the funeral conductor was clearly in charge of the funeral, with his/her way of doing things predominating for the task. This conductor would collect the funeral card, which contained the funeral specifications, and keep it in his/her possession until the interment was completed.

In the coordination office, however, the staff shared tasks. Because Dover Funeral Services worked on a roster system, it was usual for a staff member to commence a job that would be finished by some one else. Hence the coordination staff used notes and diaries to communicate with people on other shifts, and use the same files as references.

The Physical Environment

The major difference in the physical environment between the coordination office and the funeral staff was the basic work-space. The coordination office staff actually had an office allocated to them. The funeral staff did not. It was interesting to note that this was reflected in the terminology used in the organization. The coordination office staff were referred to as "the office" or "the coordination office people" etc. The funeral staff were never known as the "funeral office" or the "funeral department", they were always referred to as the "funeral staff" or "the conductors".

While the funeral staff did not have their own work space, they did have their own arranger bags, which were not lent to other staff. These bags were labelled with the arranger's visiting card and each arranger managed the contents of these bags according to their own preferences. The coordination staff, as mentioned before, shared space, files and furniture. The funeral staff did not share space so much as "squat" where convenient. The funeral card, their major documentation, was possessed and used by one person, who was in charge of the task of conducting funerals.

Discussion

It seems clear that a relationship can be discerned between the physical artefacts reported here and the culture of the research site.

The fact that the managers used the same physical space as the staff may have contributed to the cordial relations between them. However the possession of their own physical space by managers in the form of their own offices may, at the same time, have emphasised their higher position in the organization. Thus, the physical artefacts reflected the position of the managers and the appropriate interactions with staff (Giddens, 1984; Kingery, 1993; Oldham & Rotchford, 1983; Proshansky et al., 1995; Rafaeli, Dutton, Harquail, & Mackie-Lewis, 1997; Rosen et al., 1990).

The client service principle was demonstrated in the organization's physical place. The client areas were differently presented to the functional areas. Clearly more money was spent on them to create an ambience of well-bred comfort. Far less money was spent on the functional areas; floor coverings were shabby and the paint was peeling, however, newspapers were bought for staff and bathrooms were provided, indicating that the physical environment was conducive to providing the service in comfort, but without unnecessary frills.

Interestingly, when staff wore the car-washing gear, they rarely set foot past the kitchen area during the time the researcher was there, even though the client area was several rooms away. It was also interesting to note that staff were rarely seen in the front of the building, in the client areas, unless they were actually dealing with clients, thus the client areas were seldom used for meetings by staff. This is perhaps similar to Gagliardi's (1990) example of physical space being associated with organizational norms. In the case of Dover Funeral Services it was considered inappropriate to use certain areas except when dealing face-to-face with clients.

Differences between the coordination area and the funeral staff had a clear relationship to how space was allocated and used. The individuality of the funeral staff was symbolically represented in the arranger bag. The funeral staff member could be thought of as having only a few spaces and artefacts that were his or hers, such as the bag and their individual pigeon-hole, but these places were **indisputably** his or hers. There was no pressure to share as with the coordination staff. If there was conflict with another member of the funeral staff, the staff member could walk away and do their work elsewhere. When conducting a funeral, one person was in charge; if the two other funeral staff did not agree with the way the funeral was conducted, they had to put up with it.

The coordination staff, however, were tied by their duties to a specific physical space that they were obliged to share with other people. Collaboration and sharing of resources was necessary because they relied on each other to perform and finish tasks that they had started. They also tended to be identified as a group, and if one person made a mistake, they were all implicated. As in Schein's (1985) example of Multi and Action, the physical layout clearly reflected group norms and ways of working for both the coordination staff and the funeral staff. It should be noted that the lack of an office or shared work space for the funeral staff was not mentioned or alluded to in anyway by any member of the organization. The researcher had to ask a direct question before the information concerning the lack of work space for funeral staff was forthcoming. This would indicate that the organizational members were simply not conscious of anything unusual. The work space was very much in the background of their minds (Proshansky et al., 1995).

CONCLUSIONS

The primary purpose of this paper is to explore the relationship between an organization's internal physical environment and the complex web of culture, structure and routines that make up the workplace. It is argued that with regard to Gidden's model of structuration, that the physical environment and artefacts pertaining to that environment form part of the social system that influences and is influenced by human agency. This paper has shown, through the literature and by empirical research, how the physical environment is caught within the web of role associations, symbolic meaning and hierarchical and power relationships that influence organizations. As a new member joins the organization, the physical environment reinforces the processes of socialization that the new member undergoes. The field study did not show how the physical environment may influence the organizational learning that contributes to the achievement of strategic goals, and, as implied, there is clearly a gap in the literature that could be filled by further research.

This paper also implies a need for caution on the part of organizational practitioners who implement organizational interventions. The relationship between the organizational constructs such as structure, culture and work performance, and the physical environment would indicate that such interventions need to take into account a complex set of variables before they could hope to succeed.

Physical layout and artefacts, despite being taken largely for granted in daily living, continually embody meanings and representations that are important to the people whom they surround. The implication here is that any exploration of culture or organizational life that does not include some degree of symbolic interpretation of the corporeal setting is omitting an important, though complex, store of meaning.

REFERENCES

Argyris, C. (1992/1999). *On Organizational Learning* (2nd ed.). Oxford: Blackwell Business.

Assmann, J. (1995). Collective Memory and Cultural Identity. *New German Critique, 65*, 125 -133.

Brenner, P., & Cornell, P. (1994). *The Balance Between Privacy and Collaboration in Knowledge Worker Teams.* Paper presented at the Human Factors in Organizational Design and Management., Stockholm, Sweden.

Cook, S. D., & Yanow, D. (1996). Culture and Organizational Learning. In M. D. Cohen & L. S. Sproull (Eds.), *Organizational Learning* (pp. 430-459). Thousand Oaks, CA: Sage.

Czarniawska, B. (1997). Learning Organizing in a Changing Institutional Order - Examples from city Management in Warsaw. *Management Learning, 28*(4), 475 - 495.

de Geus, A. (1988). Planning as Learning. *Harvard Business Review, 88*(2), 70 - 74.

Domingues, J. M. (1997). Social memory, social creativity and collective subjectivity. *Social Science Information, 36*(3), 469 - 492.

Doxtater, D. (1990). Meaning of the Workplace: Using Ideas of Ritual Space in Design. In P. Gagliardi (Ed.), *Symbols and Artifacts: Views of the Corporate Landscape* (pp. 107 - 128). Berlin: Walter de Gruyter.

Field, L. (1995). *Managing Organizational Learning: From Rhetoric to Reality*. Melbourne: Longman.

Fischer, G.-N. (1997). *Individuals and Environment: A Psychosocial Approach to Workspace* (Atkin-Etienne, Ruth, Trans.). Berlin: Walter de Gruyter.

Gagliardi, P. (1990). Artifacts as Pathways and Remains of Organizational Life. In P. Gagliardi (Ed.), *Symbols and Artifacts: Views of the Corporate Landscape* (pp. 3 - 38). Berlin: Walter de Gruyter.

Gagliardi, P. (1996). Exploring the Aesthetic Side of Organizational Life. In S. Clegg, C. Hardy, & W. R. Nord (Eds.), *Handbook of Organizational Studies* . London: Sage.

Giddens, A. (1984). *The Constitution of Society*. Cambridge: Polity Press.

Gieryn, T. F. (2000). A Space for Place in Sociology. *Annual Review of Sociology, 26*, 463 - 496.

Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and Qualitative Design in Educational Research*. Florida: Academic Press Inc.

Hatch, M. J. (1990). The Symbolics of Office Design: An Empirical Exploration. In P. Gagliardi (Ed.), *Symbols and Artifacts: Views of the Corporate Landscape* (pp. 129 - 146). Berlin: Walter de Gruyter.

Huber, G. P. (1996). Organizational Learning: The Contributing Processes and the Literature. In M. D. Cohen & L. S. Sproull (Eds.), *Organizational Learning* (pp. 124 - 162). Thousand Oaks, CA: Sage.

Kingery, W. D. (1993). Technological Systems and Some Implications with Regard to Continuity and Change. In S. Lubar & W. D. Kingery (Eds.), *History from Things: Essays on Material Culture* (pp. 215 - 230). Washington: Smithsonian Institute Press.

Larsen, J., & Schultz, M. (1990). Artifacts in a Bureaucratic Monastery. In P. Gagliardi (Ed.), *Symbols and Artifacts: Views of the Corporate Landscape* (pp. 281 - 302). Berlin: Walter de Gruyter.

Larson-Knight, B. (2000). Leadership, Culture and Organizational Learning. In K. Leithwood (Ed.), *Understanding Schools as Intelligence Systems* (Vol. 4, pp. 125 - 140). Stamford,: JAI Press Inc.

O'Toole, K. M. (Paddy). (2000a). The Nature of Organizational Memory, *Knowledge Management: Concepts and Controversies* . University of Warwick.

O'Toole, K. M. (Paddy). (2000b). *Organizational Memory: The Structure and Processes of Collective Remembering*. Unpublished MA dissertation, Flinders University of South Australia, Adelaide.

Oldham, G. R., & Rotchford, N. L. (1983). Relationships between Office Characteristics and Employee Reactions: A Study of the Physical Environment. *Administrative Science Quarterly, 28*, 542 - 556.

Pedler, M., Burgoyne, J., & Boydell, T. (1991/1997). *The Learning Company: A Strategy for Sustainable Development* (2nd ed.). London: McGraw-Hill.

Proshansky, H. M., Fabian, A. k., & Kaminoff, R. (1995). Place-identity: physical world socialization of the self. In L. Groat (Ed.), *Giving Places Meaning* (pp. 87 - 114). London: Harcourt Brace & Company.

Rafaeli, A., Dutton, J., Harquail, C. V., & Mackie-Lewis, S. (1997). Navigating by Attire: The Use of Dress by Female Administrative Employees. *Academy of Management Journal, 40*(1), 9 - 45.

Rosen, M., Orlikowski, W. J., & Schmahmann, K. S. (1990). Building Buildings and Living Lives: A Critique of Bureaucracy, Ideology and Concrete Artifacts. In P. Gagliardi (Ed.), *Symbols and Artifacts: Views of the Corporate Landscape* (pp. 69 - 84). Berlin: Walter de Gruyter.

Schein, E. (1985). *Organizational Culture and Leadership*. San Francisco: Jossey-Bass.

Strati, A. (1990). Aesthetics and Organizational Skills. In B. A. Turner (Ed.), *Organizational Symbolism* . New York: Walter de Gruyter.

von Krogh, G., Ichijo, K., & Nonaka, I. (2000). *Enabling Knowledge Creation: How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation*. Oxford: Oxford University Press.

Witkin, R. W. (1990). The Aesthetic Imperative of a Rational-Technical Machinery: A Study in Organizational Control Through the Design of Artifacts. In P. Gagliardi (Ed.), *Symbols and Artifacts: Views of the Corporate Landscape* (pp. 325 - 338). Berlin: Walter de Gruyter.

The Development of Scales to Measure Students' Teachers' and Scientists' Views on STS

Debra K Tedman
Flinders University of South Australia *dtedman@stjohns.sa.edu.au*

John P Keeves
Flinders University of South Australia *john.keeves@flinders.edu.au*

The starting point for this work on the development of scales was an existing instrument concerned with Views on Science, Technology and Society (STS) which had been prepared in Canada. This Australian study developed scales to measure views towards Science, Technology and Society, and, it was necessary initially to specify scores to the alternative responses or views for each of the statements included in the scales used in this study. The initial scores or codes for the scales were based upon preliminary analysis and the researcher's judgment derived from a review of the literature. Subsequently, a validation study used the opinions of experts to confirm the numerical codes assigned to the responses. It was also necessary to test the items in each of the scales to see whether the model of a unidimensional scale was consistent with recorded data. It was possible to show that by using the numerical codes, the chosen items fitted well their respective scales. Once the three scales (a) effects of Society on Science and Technology (Society), (b) the effects of Science and Technology upon Society (Science), and (c) characteristics of Scientists (Scientists), had been specified and items were identified that satisfied the requirement of unidimensionality, it was possible to calibrate the three scales and the items within them using the partial credit model for Rasch scaling. The construction and calibration of these three scales permitted an investigation to proceed that involved the accurate measurement of students', teachers' and scientists' views on STS.

Science, technology and society, Science education, Australia, Rasch scaling, scales, validation

PROBLEM

One of the aims of this Australian study was to develop a master scale to measure views, beliefs and attitudes towards Science, Technology and Society (STS). Oppenheim (1992) has stressed very strongly that a great deal of careful questioning, thought and consideration, and repeated conceptualisations, are necessary to produce effective attitude scales. Furthermore, he has suggested that after trialling in a pilot study a large number of attitude statements, this item pool should be analysed and submitted to a scaling procedure. The resultant scales, each of which would consist of a smaller number of statements than in the original item pool, could be used to allocate a numerical score for each respondent. A procedure similar to the one described by Oppenheim for the development of attitude scales was used in this study, although the study sought to construct what may best be considered as 'view' or 'descriptive' scales (Morgenstern and Keeves, 1997) rather than attitude scales.

The development of the scales, the calibration of the scales and the determination of how well the data from the respondents fit the requirements of Rasch scaling are discussed in this paper. Moreover, the validation of the scales by using the opinions of the experts is considered.

THE ORIGINAL VOSTS INSTRUMENT

The scales used in this Australian study were constructed and calibrated from the *Views on Science-Technology-Society (VOSTS)* instrument (Aikenhead, Fleming and Ryan, 1987) originally developed in Canada. The three scales finally used in this Australian study related to:

(1) effects of Society on Science and Technology (*Society*);

(2) effects of Science and Technology upon Society (*Science*); and

(3) characteristics of Scientists (*Scientists*).

Each item used in the scales addressed a particular issue. The item and the issue addressed by each particular item are shown in Appendix 1.

Aikenhead and Ryan (1992) believed that the processes used to write the VOSTS items from students' perspectives conferred an inherent validity to these items that were prepared from students' actual statements. They believed that it was, therefore, not appropriate to speak in the traditional sense about the validity of VOSTS items. Moreover, Aikenhead and Ryan concluded that "the field of Item Response Theory has not yet developed the mathematical procedure that could analyse responses to VOSTS items" (Aikenhead and Ryan, 1992, p. 488).

It must be argued, nevertheless, that Aikenhead and Ryan were apparently unaware of the ways in which the partial credit model of item response theory might be employed to calibrate the students' responses to the items that they had developed, provided the items and the assigned code values were consistent with the requirement of unidimensionality associated with the scale domain that they had defined. In order to measure students' views and to compare their views with those of their teachers and professional scientists in a systematic way this present study challenged Aikenhead and Ryan's contention. In addition, this study sought to show that there was internal consistency within both the categories of items that Aikenhead and Ryan had assembled, and within the underlying views of students in relation to STS issues that provided strong meaning to their responses. Thus, the validation of the instrument was a necessary and critical component of this use of scale scores to compare the views of students, teachers and scientists, alongside the use of calibration procedures based on item response theory and the Rasch model.

MODELS OF MEASUREMENT

For the purposes of this Australian study, the form of statistical analysis used to establish unidimensionality and consequent interval scale measurement was the Rasch partial credit model (Rasch, 1960) based on the item response theory analysis of the students' levels of performance in relation to the items of the instrument. This technique employs a latent trait model, and a theoretical approach to educational measurement with items in which more than two ordered levels of outcome are defined (Masters, 1988). This models has considerable advantages when compared with the classical test theory models. The characteristics and advantages of this model are discussed below, following a discussion of the limitations of classical test theory models.

Item response theory

The measurement processes in this Australian study employ a scaling model in which scale values are assigned to both the items and the respondents. As Keeves (1992) wrote in regard to item response theory, it is a requirement that there is a common underlying trait of performance for both respondents and the items which are used. The items and the respondents are located at levels along a scale defined by the latent trait. In this study, the latent trait is the strength and coherence of students' views towards issues resulting from the relationships between science, technology and society. The position of an individual respondent on the latent trait scale is the level at which the respondent would answer, with a specified degree of probability (usually 0.50), an item located at that level on the scale (Keeves, 1992). In the scaling model used in this study, the respondents depend on the strength and coherence of their underlying views towards STS to respond to the items in a favourable manner.

The model of measurement used in this study is the partial credit model of item response theory. The two basic postulates of item response theory are:

1. the performance of an examinee on items can be predicted by factors called latent traits; and
2. the relationship between item performance and the set of underlying traits can be described by a monotonically increasing curve called an 'item characteristic curve' (Hambleton, Swaminathan and Rogers, 1991; Weiss and Yoes, 1991).

This item characteristic curve relates the probability of success on an item to the pereformance measured by the test and the characteristics of the item (Hambleton, Swaminathan and Rogers, 1991). The particular item response model used in this study is the Rasch partial credit model. The Rasch model is a one-parameter logistic model which has strong measurement properties. The advantage of item response theory over classical test theory is that the scores provided are not dependent on the specific set of items administered or the specific group of persons used in calibration (Weiss and Yoes, 1991).

In order to use the Rasch partial credit model, it is necessary to assume that the items occupy a space along the latent trait continuum that is consistent across both the entire group of students and the set of items being used (Keeves, 1988). Thus the partial credit model is a measurement model, since it provides a probabilistic connection between the categories of observed outcome on an item and the respondents' location on a latent trait of developing views. This probabilistic connection provides a basis for constructing measures of the respondent's views from a set of items with multiple outcome categories (Masters, 1988). Aikenhead and Ryan (1992) had not realised that item response theory had advanced to the point of encompassing several levels of student response. Hence this study, which uses item response theory to establish the consistency of the scaling of the VOSTS items developed by Aikenhead and his colleagues (Aikenhead, Fleming and Ryan, 1987) is a significant advance in research in this area.

Issues of consistency and validity

The first step in scaling the VOSTS items involved determining response values, or establishing which of the responses for each item should be allocated the chosen range of codes of 4, 3, 2, 1, or 0. This initial scaling was performed after consideration of the scholarly writing on science, technology and society. It was then important to examine the internal consistency of each scale using item response theory, and to establish the validity of each scale using the views of a panel of experts.

Through the use of the QUEST program (Adams and Khoo, 1993), the values for strength and coherence of the views of the respondents corresponding to each score level were determined and estimates were obtained for the difficulty parameters of each of the items in the scales. It was then possible to determine how well the Rasch model fitted the data. This was necessary, since as Wright (1988) has stressed, if the model did not fit the data, then it could not be used to calibrate items or to measure persons. Thus, in Wright's opinion, it was necessary for measurement to examine the validity of both item response patterns, and person response patterns, by evaluating the fit between the model and the data.

Response validation by analysis of fit between the model and the data

The codes for the different statements that comprised the items were specified and the location of each item estimated indicate its position along the graph of the latent trait generally, with the high codes at the top of the graph and the lower codes towards the bottom. The map would also show the position of the response groups that were formed on the same scale. It was expected that statements for each item with equivalent code values would be located at similar threshold levels along the latent trait axis. Moreover, if the codes for an item differed substantially in position from that of the equivalent codes for other items, either the item should be rejected, or the assigned codes should be reconsidered and adjusted. The scales used in the study could, in this way, be calibrated for use. If the fit of an item to the scale had an acceptable value, the item was considered to be consistent with the underlying scale.

The consistency of each scale was tested by examining the data gathered during a pilot study using a group of respondents with similar characteristics to those for whom the scales were intended (Hambleton and Zaal, 1991). In this study, the consistencies of the items and statements and the students' responses were initially examined before the scales were employed to measure the views of senior secondary science students in the 29 South Australian schools and colleges which were sampled for the main study. The internal consistency was therefore demonstrated, in the first instance, by how well the data fitted the Rasch scale. In this study, the analysis also demonstrated the degree to which the items satisfied the requirement of unidimensionality, which reflected the validity of the scale.

Furthermore, invariance of item and performance parameters, which only holds when the data fit the model well, is an indispensable feature of Rasch measurement, since it enables applications of the modelling procedures in such investigations as those of item bias and readily permits both vertical and horizontal equating (Hambleton, Swaminathan and Rogers, 1991). Therefore, it is very important to determine whether the partial credit model fitted the data in this study. The fit of the models to the student data was sufficiently strong for it to be argued that the scales were well-supported, and that the researcher could proceed confidently with the investigation and further data analysis.

Internal consistency and validity in item response theory

In evaluating the instrument used in this Australian study, it was important to address the issues of validity as well as internal consistency and reliability, in order to determine the generalisability of the scores allocated to respondents, and the strength of measurement that could be made using the scores. After the internal consistency of the items was examined using Rasch scaling the validity was considered. One conception of validity is how faithfully the set of items in an instrument correspond to that attribute in which the researchers are interested (Hambleton and Rogers, 1991). The scores are valid if, in fact, the instrument is seen to measure what it purports to measure.

In order to test the validity of each scale seven persons who were recognised scholars in the field of Science, Technology and Society (STS) were asked to rank the statements in each item on a five-point scale according to their understanding of the underlying STS dimension. These aggregated rankings could then be compared with the scale scores assigned by the researchers from a systematic review of the literature on STS.

Table 1 for Item No. 3 shows the results for the assigned scores (Scale scores) for this item in comparison with the experts' responses that are recorded as mean scores (Expert mean scores), ranks (Expert ranks) and integral scale scores (Expert scores). After the experts' scores were collected, the items were separated by scale, and three tables were prepared, with the items in presentation order. The agreement between the experts' scores and the final scores used in this present study could then be examined. First, the experts' scores were pooled and the experts' mean scores for the responses for each item were calculated. The experts' mean scores were then ranked. Overall experts' scores were subsequently determined according to this ranking.

Table 1. Validation of response scores using the ratings of the experts - Scale: Society

Item No. 3									
Response Category	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Expert mean scores	2.00	2.43	3.14	3.00	1.43	1.86	0	0	0
Expert ranks	4	3	1	2	6	5	7	7	7
Expert scores	2	3	4	4	1	2	0	0	0
Scale scores	2	2	3	4	1	1	0	0	0
χ^2_7	**p [a]=**	**0.14**		**α=0.94**		**% ag [c]=**	**0.57**		

a: probability value; b: reliability coefficient; c: percentage agreement

Secondly, the level of agreement, and significance of any differences between the researcher's scale scores and overall experts' scores were examined using χ^2 values and percentage agreement. For each item, the χ^2 value and its probability were calculated using the reliability analysis routine in SPSS (Version 7) with the Friedman ANOVA option, with zero forming only one category. The results for each scale are given in Table 2.

The reliability coefficient was obtained as a Cronbach α coefficient also from the SPSS program. The percentage agreement was calculated for each item by dividing the number of responses where there was complete agreement between the researcher's scale scores and the experts' scale scores by the total number of responses for the item. For Item 3, by way of example, the percentage agreement (0.57) is calculated by dividing the number of responses which are in agreement (4) by the total number of responses (7). Where there is a very high level of agreement, the reliability coefficient goes to one, and the per cent agreement goes to one.

The coefficients of concordance were not employed to indicate the significance of any difference between the scores and the experts' score, since they could not be used for all cases. However, the method of validation using χ^2 values, reliability coefficients and percentage agreement showed that while the experts' scores and those of the researcher were not in complete agreement, there was sufficient agreement to indicate that the scoring employed was valid or strong in so far as it was in agreement with the scoring provided by the experts.

Table 2. Summary statistics for the validation of response scores using the ratings of the experts for Society, Science and Scientists Scales

Item No.	χ^2	p [a]	R C [b]	% ag [c]
Society Scale				
3	$\chi^2_7 = 4.0$	0.14	0.94	0.57
5	$\chi^2_9 = 3.0$	0.39	0.92	0.44
9	zero		1.00	1.00
12	$\chi^2_9 = 3.2$	0.20	0.93	0.67
15	zero		0.98	0.86
19	zero		1.00	1.00
21	$\chi^2_6 = 0.0$	1.0	0.93	0.50
23	zero		1.00	0.67
25	zero		1.00	0.33
Science Scale				
1	$\chi^2_8 = 4.4$	0.35	0.85	0.38
2	$\chi^2_6 = 3.2$	0.16	0.97	0.67
7	zero		0.98	0.86
10	$\chi^2_9 = 2.0$	0.16	0.97	0.78
14	zero		1.00	1.00
16	$\chi^2_8 = 2.0$	0.37	0.93	0.63
17	zero		1.00	1.00
22	zero		1.00	1.00
24	zero		0.95	0.71
26	zero		0.89	0.43
Scientists Scale				
4	zero		0.99	0.90
6	zero		0.95	0.90
8	zero		0.95	0.60
11	$\chi^2_9 = 3.2$	0.20	0.94	0.50
13	$\chi^2_9 = 2.0$	0.16	0.97	0.78
18	$\chi^2_7 = 5.8$	0.21	0.89	0.28
20	$\chi^2_{10} = 5.3$	0.38	0.89	0.40
27	$\chi^2_9 = 4.4$	0.22	0.94	0.56

a: probability value; b: reliability coefficient; c: percentage agreement

The Calibration of the Scales

During this study, analysis of the fit of a unidimensional model to the data was conducted by using Rasch scaling to place the people and the items on the same unidimensional interval scale. The fit between the model and the data was established primarily by using the item infit mean square values. This analysis also examined the item analysis results obtained with classical test theory, item difficulty level estimates or threshold values and maps of respondents and items for each scale. Hence, it is important at this stage to describe, in some detail, the techniques of analysis which were employed.

In item response theory, the probability of a correct response depends on both the respondent's performance level and on the difficulty parameter of the item, which is its difficulty level or threshold value. In this study, both the respondents' performances and the item threshold parameters were initially unknown. After the data had been collected and entered, the performance parameters of respondents and the Thurstone threshold values for the response categories for each item and the fit parameters of the items and the persons were estimated using the QUEST computer program (Adams and Khoo, 1993). The parameters that were associated with the theoretical item response curve were estimated using a maximum likelihood procedure.

A minor disadvantage of this procedure was that performance estimates associated with perfect and zero scores did not provide information for the analysis. There were no zero scores in the sample, so these did not need to be considered. It was necessary, however to exclude those students obtaining perfect scores in the main analysis and subsequently to calculate the estimates for their perfect scores. The units on this scale are logits (Hambleton and Swaminathan, 1985). The following section considers the information provided in the QUEST program for the analysis of a particular item. Appendix 2 records the items included in the three scales.

Item Analysis and Response Values

Table 3 shows the item analysis results for observed responses to Item 3 for students for the three scales. There are nine possible multiple-choice responses for Item 3, and ten categories of response: 0,1,2,3,4,5,6,7,8, and 9 if the omitted response is also considered. The QUEST computer program (Adams and Khoo, 1993) was used for item analysis in this study. The first task in the development of a scoring procedure for the VOSTS instrument involved the recoding from the original alphabetical or numerical coding system for the responses for each item to item characteristic scores, between zero and four, which were allocated to the responses for particular items. Hence, these numerical categories of response were recoded to 0,2,2,3,4,1,1,0,0, and 0 once scores were allocated to each response. This formed a five-point scale, since scores from 0-4 were used. The initial scores were assigned by the researcher based on an understanding of STS theoretical positions. Those response categories most consistent with the 'ideal' STS position, were assigned a score of 4 and those least consistent with the 'ideal' STS position were assigned a score of 0, with intermediate scores assigned accordingly. The item analysis information for each of the categories includes: counts, percentages, point-biserial correlations, p-values, mean ability levels, step labels, Thurstone thresholds and errors.

This procedure was similar for all of the other items, and the results for the observed responses demanded very careful examination during this stage of the study. A substantial amount of very important information was provided by the distractors or alternative response categories that Aikenhead and his colleagues developed, and it was important to use this information, rather than to ignore it.

The point-biserial values (point-biserial correlations) provide information on the discrimination for each response statement (Hambleton and Swaminathan, 1985). For Item 3 (see Table 3), the point-biserials related well with the mean performance values for students responding to this response statement.

Table 3. Item Analysis results for observed responses: students on Item 3 on Society Scale (N = 1278; L = 9)

Item 3 (Disc[a] = 0.47) Infit MNSQ[b] = 1.06

Categ[c]	0	1	2	3	4	5	6	7	8	9
Score Code	0	2	2	3	4	1	1	0	0	0
Counts[d]	3	203	109	147	382	277	83	14	32	28
Percent[e]	0.2	15.9	8.5	11.5	29.9	21.7	6.5	1.1	2.5	2.2
Pt-Biser[f]	-0.02	0.01	-0.07	0.07	0.37	-0.13	-0.17	-0.21	-0.23	-0.15
p-value[g]	.221	.425	.005	.004	.000	.000	.000	.000	.000	.000
Mean Ability[h]	0.02	0.23	0.11	0.35	0.53	0.10	-0.10	-0.84	-0.61	-0.30
Step Labels		1	2	3	4					
Threshold[i]		-1.75	-0.13	0.41	0.73					
Error		0.16	0.09	0.09	0.10					

a: discrimination index; b: infit mean square; c: response category d: number of respondents; e: percentage of respondents; f: point-biserial correlations; g: probability that the point biserial correlations are significantly different from zero; h: mean ability level for respondents; i: Thurstone thresholds or difficulty levels

By way of example, response Category 4, which drew 382 (29.9%) students has been assigned the highest score of four and has the highest mean ability level for respondents of 0.53 with a point-biserial correlation of 0.37. The next highest score is three for Category 3. The mean ability level is 0.35, resulting in the second highest mean ability level for all response categories and a point-biserial correlation of 0.07. Some of the lowest point-biserial values are for categories 0, 7, 8, and 9 with negative point-biserial correlations of -0.02, -0.21, -0.23, and -0.15 respectively. These categories also have low mean respondent ability levels of 0.02, -0.84, -0.61, and -0.30 respectively.

The p-values merely indicate the probability that the point-biserial correlations are significantly different from zero. The mean ability values in the item analysis also provide important information since the less favourable the response category towards the 'ideal' STS position, the lower the mean ability value. The mean ability values indicate the relative degree of strength and coherence of a response category, with the stronger and most coherent response categories having the highest mean ability values. Hence, the mean ability levels for Category 4 (0.53) and Category 3 (0.35) indicate response categories closest to the 'ideal' STS position.

In the pilot study, in those few cases where discrepancies were observed, some adjustments were made since it was considered important to re-consider the scaling. The students were found, overall, to have quite strong, coherent and well-directed views towards STS. The review of the literature was taken into account during this process of further reflection upon the scores allocated to each of the responses.

The overall consistency between the scaled response categories and the total scores of the students on the scale is given by the item discrimination index. In general, these indices are moderate or strong and the value of 0.47 is recorded for Item 3. Table 3 shows that for Item 3, the calculated value of the infit mean square deviated only slightly (1.06) from the expected value of 1.00.

Item Threshold Estimates

Table 4 presents the item estimates and threshold values for students' responses to the Society items scale showing the thresholds (between 1 and 4), and the infit and outfit mean square values and t-values for each item. Table 4 also shows the score for the item using score values (0 to 4) and the maximum overall score. The outfit mean square values give the overall fit of the item to the scale across all positions on the scale, and all persons in the sample. The infit mean square value is an index of item fit at the steepest part of the item characteristic curve. The fit of the item to the unidimensional scale is thus given by the infit mean square, with an expected value of 1.00 for an item that discriminates well. The infit mean square values should lie within the range of 0.83 to 1.20 for the item to be considered to fit the model well. In the analysis of the trial data, one item with infit mean square value of 0.80 was found to discriminate too well, and another item with an infit mean square value of 1.22 did not discriminate well enough.

The item threshold statistics show the difficulty levels associated with the response categories of an item (Masters, 1988). Table 4 shows that for Item 3, the transition levels at the 0.5 probability level of response from the score of 0 to 1, from 1 to 2, from 2 to 3, and from 3 to 4 are given by the threshold levels of -1.75, -0.13, 0.41, and 0.73 respectively. These threshold values form a clear monotonic sequence with relatively small errors (approximately 0.10) involved in estimation. The threshold for a score of four is 0.73, which is higher than the threshold of 0.41 for a score of three (see Table 4). The threshold for a score of one is -1.75, that is substantially lower than the threshold for the higher scores. Consequently, the fit estimates for this item conform well with the requirements of the model. The infit and outfit mean square values and t-values for this item are also within the range (approximately one) which indicate that the item discriminates consistently with respect to the other items in the scale.

However, the outfit mean square value is primarily used to assess the fit of persons to the model, and the t-statistics that are heavily influenced by sample size and in this study with 1278 students, are not taken into consideration. Nevertheless, it should be noted that for this scale both the infit and outfit t-values are less than the critical level of 2.0 for non-fit of an item to the scale. The fit data for all the items are presented in Appendix 3.

Table 4. Item estimates (thresholds) in input order: students on Society Scale (N = 1278; L = 9)

Item Name	Score	Max[a]	Thurstone Thresholds				Infft[b] Mnsq	Outft[c] Mnsq	Inft[d] t	Outft[e] t
			1	2	3	4				
Item 3	2949	5112	-1.75	-0.13	0.41	0.73	1.06	1.08	1.8	1.8
			.16	.09	.09	.10				
Item 5	2613	5112	-2.03	-0.33	0.80	1.33	1.05	1.07	1.3	1.5
			.16	.11	.11	.10				
Item 9	3125	5112	-1.09	-0.45	0.37	0.67	1.02	1.02	0.6	0.4
			.11	.12	.09	.10				
Item 12	2819	5112	-1.08	-0.37	0.43	1.35	1.05	1.05	1.4	1.0
			.11	.10	.09	.11				
Item 15	2829	5112	-0.47	-0.17	0.34	0.73	1.00	0.97	-0.1	-0.5
			.09	.09	.10	.10				
Item 19	3213	5112	-0.56	-0.24	-0.07	0.85	0.94	0.90	-1.7	-1.8
			.11	.11	.10	.09				
Item 21	2943	5112	-0.97	-0.20	0.38	0.74	0.96	0.95	-1.3	-1.0
			.09	.10	.08	.09				
Item 23	3115	5112	-0.48	-0.25	0.22	0.49	0.98	1.01	-0.7	0.2
			.09	.09	.09	.08				
Item 25	2543	5112	-0.38	-0.10	0.57	0.83	1.03	1.03	0.8	0.6
			.09	.09	.08	.11				
Mean			0.00				1.01	1.03	0.2	0.2
SD			0.13				0.04	0.06	1.2	1.2

a: maximum score b: infit mean square values c: outfit mean square values d: infit t values e: outfit t values

Calculating Perfect Scale Scores

Table 5 shows the score equivalents for all score values for the *Society* scale. Since the computer program did not calculate perfect scale scores, it was necessary to estimate the score for students on the three scales who gained a maximum score of 100 per cent. This was accomplished by consulting the score equivalence tables for the three scales and using the top three logits for calculations.

Table 5. Score equivalence table: all scores on Society Scale (N = 1426; L = 9)

Score	Estimate (logits)	Error
max = 36		
35	2.51	0.91
34	1.93	0.64
33	1.60	0.51
32	1.37	0.45
31	1.19	0.39
30	1.04	0.36
29	0.92	0.34
28	0.81	0.32
27	0.71	0.31
26	0.62	0.29
25	0.53	0.28
24	0.46	0.28
23	0.38	0.27
22	0.31	0.27
21	0.24	0.26
20	0.17	0.26
19	0.10	0.26
18	0.04	0.26
17	-0.03	0.26
16	-0.10	0.26
15	-0.17	0.26
14	-0.24	0.27
13	-0.31	0.27
12	-0.39	0.28
11	-0.47	0.29
10	-0.56	0.30
9	-0.65	0.32
8	-0.76	0.33
7	-0.88	0.36
6	-1.02	0.38
5	-1.19	0.43
4	-1.40	0.48
3	-1.67	0.56
2	-2.08	0.71
1	-2.78	0.99
36 (estimate)	**3.34**	
0 (estimate)	**-3.77**	

Table 5 shows that the maximum score is 36. The logits for scores of 33, 34, and 35 are 1.60, 1.93 and 2.51 respectively. The maximum logit is calculated by adding to the logit for highest score (35) shown on the score equivalence table (2.51) the difference between 2.51 and 1.93 (0.58) and the difference between 0.58 and 0.33 (1.93 minus 1.60). The figure obtained from subtracting 0.33 from 0.58 is 0.25. Thus, the maximum logit = 2.51+ 0.58+ 0.25= 3.34.

Maps of Respondents and Item Thresholds

The map of respondents and item thresholds for the *Society* scale presented in Figure 1 shows the location of the item thresholds and the respondents on the scale. The clustering of the item levels provides solid evidence for the internal consistency and validity of both the model and the scores obtained as a result of the study. Both person and item locations are shown in the maps for the three scales so they are extremely useful for the subsequent analysis of observed responses. There is more than one threshold per item on these maps because it has employed the partial credit model. The locations of the item thresholds in Figure 1 show that the levels associated with thresholds for Level 4 in the main are above the levels of thresholds for Level 3, which are higher on the scale than thresholds for Level 2. This further confirms the consistency of the scores assigned to the responses to the items in the instrument used in this study. Furthermore, the map also shows that the item thresholds are not equally spaced, but are located at particular levels on the latent trait scale. Likewise the persons are located at particular levels on the latent trait scale. The higher the level of the person on the y-axis, the greater the probability that he or she will give a strong and coherent response to an item. Similarly, the higher the level of an item threshold on the y-axis, the greater the strength and coherence of this item towards STS.

The item responses and numbers of respondents for Item 3 are shown on the item estimates map for all students on the *Society* Scale. Although the response thresholds for this item are lower than for many of the other items, the thresholds are well placed in comparison with the thresholds for the other items.

COMPARISON WITH OTHER METHODS USED TO VALIDATE ATTITUDE SCALES

There are many similarities between the methods of validation used for the attitude scales developed previously by other researchers (Rubba, Bradford and Harkness, 1996) and the validation of the scales developed in this present study. Whereas in the present study, the experts' scaling followed the scaling of the items based on a review of the literature and the researchers' judgment, in the study by Rubba, Bradford and Harkness, the scaling was performed exclusively by the panellists. In this present study, techniques discussed in this paper were used to establish the agreement between the experts' scores and the scale scores allocated by the researchers on the basis of the review of the literature.

Although Rubba and his colleagues scaled the VOSTS instrument in their study, and used the views of experts to validate the scales, as has been accomplished in this Australian study, they did not examine the items and the response categories in the way that has been done in this study. These American researchers just assumed that the experts had given the correct responses. In this Australian study it was considered important to calibrate the scaled information and the procedures to establish the validity and consistency of the scales.

2.0					
	x				
	x	**5.4**	**12.4**		
	x				
	xxx				
1.0					
	xxxx				
	xxxxxx	**5.3**	**19.4**	**25.4**	
	xxxxxx	**3.4**	**15.4**	**21.4**	
	xxxxxx	**9.4**			
	xxxxxxxx	**25.3**			
	xxxxxxx	**12.3**	**23.4**		
	xxxxxxxxxxxxxxxxxxxx	**3.3**	**9.3**	**15.3**	**21.3**
	xxxxxxxx				
	xxxxxxxx	**23.3**			
	xxxxxxxx				
0.0	**xxxxxxxxxxx**				
	xxxxx	**3.2**	**19.3**	**25.2**	
	xxxx	**15.2**	**21.2**		
	xxxxxxxx	**19.2**	**23.2**		
	xxx	**5.2**	**12.2**	**25.1**	
	xx	**9.2**	**15.1**	**23.1**	
	xx	**19.1**			
	xx				
	x				
	x				
-1.0	**x**	**21.1**			
		9.1	**12.1**		
	x				
		3.1			
-2.0		**5.1**			

Each x represents 10 students
Scale zero set at mean of item thresholds

Figure 1. Map of respondents and item estimates (thresholds): all students on Society Scale (N = 1278; L = 9)

Consideration of the validity of attitude scales used in previous studies further highlights the appropriateness of the methods which were used in this present study to establish the validity of the items in the scales. The Scientific Attitude Inventory (SAI) (Moore and Sutman, 1970) was revised by Moore and Foy (1997) in order to improve readability and remove gender-biased language. There had been criticism of the validity of the original instrument (Munby, 1983). This criticism was even more pronounced in regard to the revised instrument (Munby, 1997) and was quite important and worthy of consideration. During the process of the revision of the SAI, the researchers argued that the validity of the original instrument, which was established using the judgments of a panel of judges, was maintained in the final SAII instrument, since the position of statements in the instrument had not changed. The authors wrote that:

> since there is evidence for the content validity of the items in the original instrument with respect to the 12 position statements, we decided to make as few changes as possible while responding to criticisms and suggestions. This evidence was presented in item selection and the field test of the original SAI. (Moore and Foy, 1997, p. 329)

In the original instrument, attitude statements were selected for use from a pool of items after the judges judged each attitude statement in terms of whether it represented a particular position statement (Moore and Sutman, 1970). This method of validation differs substantially from the method in the present study. Furthermore, the panel of seven experts from the Australasian Association for the History, Philosophy and Social Studies of Science provided independent scoring of the items in the scales, and the agreement between the experts' scale scores and those of the researchers was established using well-recognised statistical techniques.

The construct validity of the original SAI was demonstrated in a field test (Moore and Sutman, 1970). During this test, the SAI was administered to three groups of low-ability tenth-grade biology students. The investigator presented lessons to each of these groups. The series of lessons for the first group, the control group, was the regular sequence prepared by the teacher of this group. The other two groups received lessons which were specially designed to develop the attitudes assessed by the SAI. The authors believed that since both of the second two groups who received instruction relevant to the development of the scientific attitudes in the SAI had significantly higher post-test means than the post-test means of the control group, this field test showed that the SAI had construct validity. Munby (1983) questioned the method by which Moore and Sutman (1970) believed they had validated the SAI instrument. He argued that it was not certain what was measured by the SAI, and many of the items in the SAI, that were believed to gauge attitudes, could be interpreted very differently. Following his study of the SAI, Munby contended that "there are sufficient grounds for judging the SAI to be conceptually doubtful if not weak" (Munby, 1983, p. 157).

After their redevelopment of the SAI, Moore and Foy (1997) argued that they sought to show the validity of SAI II with a confirmatory factor analysis of data from 557 respondents. However, this resulted in a reduced number of items in five scales, and the grouping of these items was not very satisfactory in providing meanings for the item groups. The authors concluded that:

> regrouping the items would virtually eliminate the support gained by judges for the validity of the instrument. Therefore, the 40-item SAI II is being advanced as presented here without the support of factor analysis. (Moore and Foy, 1997, p. 332)

These findings cast doubt upon Moore and Foy's assertion in regard to their redeveloped SAI II scale that:

> it is possible that the objects of the scales are so ill-formed in the subjects, students in this case, that we are not able to use their responses to confirm the scales as factors. (Moore and Foy, 1987, p. 333)

Munby (1997) was rightfully very critical of Moore and Foy's claims in regard to the validity of the SAI II instrument. He showed that empirical work with the SAI II raised doubts about its validity. Munby decried the statement by Moore and Foy about the objects of the scales being so ill-formed that it was not possible to use students' responses to confirm the scales as factors. Munby argued that;

> This statement suggests that the authors put more credence in the evidence of validity obtained from the panel of judges using the older version of the SAI than they do in the empirical determination using the present version. (Munby, 1997, p. 338)

Both the empirical methods and the validation by a panel of experts which were used in the development of the scales employed in the present study have been discussed in detail. Moreover, the validation by a panel of experts differed substantially from that used by Moore and Foy. The examination of the scores assigned by the researchers and the experts employed statistical techniques to show the validity of the scoring of the scales. The need to establish the way in which the empirically-derived responses corresponded with the scales was supported by Munby (1996).

Once the fit of the student information to the scale to establish the consistency of the scoring of the scales, and the validity of the scales had been ensured, an investigator could proceed with confidence to measure respondents' views using these scales.

SUMMARY AND CONCLUSION

The scales used in this present study were developed from items in the VOSTS inventory, with careful consideration of the nature of the items used in the final instrument, as well as the consistency and the validity of the scales. As a first step in the development of the scales, each of the response categories was assigned a numerical score based on a judgment of its degree of consistency with the 'ideal' STS view. The final instrument was prepared after the examination of the data collected during the trial study.

A detailed literature review was an important form of validation of the scale values assigned to the responses for the items in the instrument. Furthermore, the establishment of the validity of the scales by comparison with the views of the experts is presented in this paper.

The trial data were used to examine whether the items discriminated sufficiently between respondents who were high on the scale and those who were low. The numbers of students who selected particular scaled responses were shown on maps for each scale, where the vertical axis represented the overall levels of strength and coherence towards STS of respondents' views. The QUEST program estimated the thresholds where one scaled response changed to another and these Thurstone thresholds have been presented. It was expected that most of the very strong and coherent scores (fours), would be at the top of the graph. This was found to occur.

Consequently, after the data for the pilot study were entered and processed, the consistency of each item to its scales was analysed, since the analysis provided evidence for a test of goodness of fit for individual items to the particular scales. It was possible to use Rasch scaling to investigate the fit of the model to the data, since the Rasch scaling model places people and items on the same interval scale. Infit mean square values were used to establish the fit between the model and the data. Items that did not conform to the model were eliminated, due to their inability to differentiate the strength and coherence of respondents' views towards STS in a manner that was consistent with the other items. This analysis of the data also examined the item analysis results for observed responses, item difficulty level estimates or threshold values and maps of respondents and items for each scale.

In conclusion, the partial credit model and Rasch measurement procedures provide a very powerful approach to the scaling of views and attitudes which sets the scores obtained on an interval scale that is not only independent of the items and statements employed, but also the sample of persons used to calibrate the scale.

REFERENCES

Adams, R.J. & Khoo, S.K. (1993). *Quest - The Interactive Test Analysis System*. Hawthorn, Victoria: Australian Council for Educational Research.

Aikenhead, G.S., Fleming, R.W. & Ryan, A.G. (1987). High school graduates' beliefs about Science-Technology-Society. 1. Methods and issues in monitoring student views. *Science Education*, 71, 145-161.

Aikenhead, G.S. & Ryan, A.G. (1992). The development of a new instrument: "Views on Science-Technology-Society" (VOSTS). *Science Education*, 76, 477-491.

Hambleton, R.K. & Rogers, H.J. (1991). Advances in criterion-referenced measurement. In *Advances in Educational and Psychological Testing*, Hambleton, R.K. & Zaal, J.N. (eds) pp. 3-413. Boston: Kluwer Academic Publishers.

Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Nijhoff Publishing.

Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, California: Sage Publications.

Keeves, J.P. (1988). *Educational Research, Methodology and Measurement: An International Handbook*. Oxford: Pergamon Press.

Keeves, J.P. (1992). *The IEA Technical Handbook*. The Hague, The Netherlands: The International Association for the Evaluation of Educational Achievement (IEA).

Masters, G.N. (1988). Partial credit models. In *Educational Research, Methodology and Measurement: An International Handbook*, Keeves, J.P. (ed). Oxford: Pergamon Press.

Moore, R.W. & Hill Foy, R.L. (1997). The Scientific Attitude Inventory: A revision (SAI II). *Journal of Research in Science Teaching*, 34, 327-336.

Moore, R.W. & Sutman, F.X. (1970). The development, field test and validation of an inventory of scientific attitudes. *Journal of Research in Science Teaching*, 7, 85-94.

Morgenstern, C. & Keeves, J.P. (1997). Descriptive Scales. In *Educational Research Methodology, and Measurement: An International Handbook*, Keeves, J.P. (ed) pp. 900-908. Cambridge, UK: Cambridge University Press.

Munby, H. (1983). Thirty studies involving the "Scientific Attitude Inventory": What confidence can we have in this instrument? *Journal of Research in Science Teaching*, 20, 141-162.

Munby, H. (1997). Issues of validity in science attitude measurement. *Journal of Research in Science Teaching*, 34, 337-341.

Oppenheim, A. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*. London: Pinter Publishers.

Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Copenhagen, Denmark: Paedagogiske Institute.

Rubba, P.A., Bradford, C.S. & Harkness, W.J. (1996). A new scoring procedure for The Views on Science-Technology-Society instrument. *International Journal of Science Education*, 18, 387-400.

Weiss, D.J. & Yoes, M.E. (1991). Item response theory. In *Advances in Educational and Psychological Testing*, Hambleton, R.K. & Zaal, J.N. (eds) pp. 69-95. Kluwer, Boston: Academic Publishers.

Wright, B.D. (1988). Rasch measurement models. In *Educational Research Methodology and Measurement: An International Handbook*, Keeves, J.P. (ed) pp. 286-297. Oxford, England: Pergamon Press.

APPENDIX 1: ISSUES ADDRESSED BY THE ITEMS IN THE SCALES

The Effects of Society on Science and Technology (Society)

The Influence of Governments on Science (Items 3,9,19, and 25)
The Influence of the Military on Science. Case History: Scientists in the Second World War (Item 23)
The Influence of Special Interest Groups on Science (Item 12)
The Influence of Educational Establishments on Science (Items 5 and 21)
The Control of Science by Corporations (Item 15)

The Effects of Science and Technology upon Society (Science)

Social Responsibility of Scientists (Item 10)
Science for the Solution of Practical Problems in Everyday Life (Item 1)
Contributions to Social Decisions in Relation to Science and Technology (Item 16)
The Contribution of Science and Technology to Economic Development (Items 2 and 22)
Unwanted Effects Caused by the Application of Science (Item 17)
The Effects of Capital-intensive Technology (Items 7 and 24).
Science, Technology and the Social Fabric of Society (Items 14 and 26)

Characteristics of Scientists

The Effect of the Personal Characteristics of Scientists on Science (Items 6 and 18)
Australian Scientists' Motivation for Doing Science (Item 27)
The Effect of the Values of Scientists on Science (Item 8)
Individual Scientists and the Scientific Method (Item 11)
Under-representation of Females in Science (Item 13)
Gender Effects on the Outcomes of Science (Items 4 and 20)

APPENDIX 2: ITEMS IN THE SCALES

Appendix 2.1 Item Number 1

In your everyday life, knowledge of science and technology helps you solve practical problems (for example, getting a car out of the sand, cooking, or caring for a pet).

Your position, basically: (Please read from A to J, and then choose ONE only from this page.)

The systematic reasoning taught in science classes (for example, hypothesising, gathering data, being logical):

A. helps me solve some problems in my daily life. Everyday problems are more easily and logically solved if treated like science problems. (2)

B. gives me greater knowledge and understanding of everyday problems. However, the problem solving techniques we learn are not directly useful in my daily life. (3)

C. Ideas and facts I learn from science classes sometimes help me solve problems or make decisions about such things as cooking, keeping healthy, or explaining a wide variety of physical events. (3)

D. The systematic reasoning and the ideas and facts I learn from science classes help me a lot. They help me solve certain problems and understand a wide variety of physical events (for example, thunder or quasars). (4)

E. What I learn from science class generally does not help me solve practical problems; but it does help me notice, relate to, and understand, the world around me. (2)

What I learn from science class does not relate to my everyday life:

F. Biology, chemistry and physics are not practical for me. They emphasise theoretical and technical details that have little to do with my day-to-day world. (1)

G. My problems are solved by past experience or by knowledge unrelated to science and technology. (2)

H. I don't understand. (0)

I. I don't know enough about this subject to make a choice. (0)

J. None of these choices fits my basic viewpoint. (0)

Scores for responses are shown in parenthesis

Appendix 2.2 Item Number 2

The more Australia's science and technology develop, the wealthier Australia will become.

Your position, basically: (Please read from A to H and then choose ONE only from this page.)

Science and technology will increase Australia's wealth:

A. because science and technology bring greater efficiency, productivity and progress. (2)

B. because more science and technology would make Australia less dependent on other countries. We could produce things for ourselves. (2)

C. because Australia could sell new ideas and technology to other countries for profit. (3)

D. It depends on which science and technologies we invest in. Some outcomes are risky. There may be other ways besides science and technology that create wealth for Australia. (4)

E. Science and technology decrease Australia's wealth because it costs a great deal of money to develop science and technology. (1)

F. I don't understand. (0)

G. I don't know enough about this subject to make a choice. (0)

H. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.3 Item Number 3

The Australian government should give scientists research money to explore the curious unknowns of nature and the universe.

Your position, basically: (Please read from A to I and then choose ONE only from this page.)

Money should be spent on scientific research:

A. so Australia does not fall behind other countries and become dependent upon them. (2)

B. in order to satisfy the human urge to know the unknown; that is, to satisfy scientific curiosity. (2)

C. even though it's often impossible to tell ahead of time whether the research will be beneficial or not. It's an investment risk but we should take it. (3)

D. because by understanding our world better scientists can make it a better place to live in (for example using nature's environment and resources to our best advantage, and by inventing helpful technology). (4)

E. only when the research is directly related to our health (especially finding cures for diseases), to our environment or to agriculture. (1)

F. Little or no money should be spent on scientific research because the money could be spent on other things such as helping Australia's unemployed and needy, or helping less fortunate countries. (1)

G. I don't understand. (0)

H. I don't know enough about this subject to make a choice. (0)

I. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.4 Item Number 4

There are many more women scientists today than there used to be. This will make a difference to the scientific discoveries which are made. Scientific discoveries made by women will tend to be different from those made by men.

Your position, basically: (Please read from A to M, and then choose ONE only from this page.)

There is NO difference between female and male scientists in the discoveries they make:

A. because any good scientist will eventually make the same discovery as another good scientist. (2)

B. because female and male scientists experience the same training. (2)

C. because overall, women and men are equally intelligent. (3)

D. because women and men are the same in terms of what they want to discover in science. (2)

E. because research goals are set by demands or desires from others besides scientists. (3)

F. because everyone is equal, no matter what they do. (1)

G. because any differences in their discoveries are due to differences between individuals. Such differences have nothing to do with being male or female. (4)

H. Women would make somewhat different discoveries because, by nature or by upbringing, females have different values, viewpoints, perspectives, or characteristics (such as sensitivity toward consequences). (4)

I. Men would make somewhat different discoveries because men are better at science than women. (1)

J. Women would likely make somewhat better discoveries than men, because women are generally better than men at some things such as instinct and memory. (1)

K. I don't understand. (0)

L. I don't know enough about this subject to make a choice. (0)

M. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.5 Item Number 5

The success of science and technology in Australia depends on us having good scientists, engineers and technicians. Therefore, Australia should require students to study more science in school.

Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

Students should be required to study more science:

A. because it is important for helping Australia to keep up with other countries. (1)

B. because science affects almost every aspect of society. As in the past, our future depends on good scientists and technologists. (3)

C. Students should be required to study more science, but a different kind of science course. Students should learn how science and technology affect their everyday lives. (4)

Students should NOT be required to study more science:

D. because other school subjects are equally or more important to Australia's successful future. (2)

E. because it won't work. Some people don't like science. If you force them to study it, it will be a waste of time and will turn people away from science. (2)

F. because not all students can understand science, even though it would help them in their life. (1)

G. because not all students can understand science. Science is not really necessary for everyone. (1)

H. because it's not right for someone else to decide if a student should take more science. (1)

I. I don't understand. (0)

J. I don't know enough about this subject to make a choice. (0)

K. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.6 Item Number 6

The best scientists have the patience and determination to get through times of frustration and boredom (for example, doing the same experiment many times to get reliable results).

Your position, basically: (Please read from A to G, and then choose ONE only from this page.)

A. Yes, because frustration and boredom challenge the best scientist to struggle and work even harder. (1)

B. Yes, because patience and determination are part of the job. Without them, scientists would not get absolutely correct results. (4)

C. No, because even some of the best scientists cannot cope with frustration. Scientists have varying degrees of patience, like everyone else. (3)

D. No, because the best scientists are clever enough to avoid most frustration and boredom. Frustration and boredom make it harder for anyone to succeed. (2)

E. I don't understand. (0)

F. I don't know enough about this subject to make a choice. (0)

G. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.7 Item Number 7

Heavy industry has greatly polluted Europe and North America. Therefore, it is a responsible decision to move heavy industry to underdeveloped countries where pollution is not so widespread.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

A. Heavy industry should be moved to underdeveloped countries to save developed countries and their future generations from pollution. (1)

B. It's hard to tell. By moving industry, developed countries would help poor countries to prosper, and developed countries would help reduce their own pollution. But they have no right to pollute someone else's environment. (2)

C. It doesn't matter where industry is located. The effects of pollution are global. (3)

Heavy industry should NOT be moved to underdeveloped countries:

D. because moving industry is not a responsible way of solving pollution. Developed countries should reduce or eliminate their own pollution, rather than create more problems elsewhere. (4)

E. because those countries have enough problems without the added problem of pollution. (2)

F. because pollution should be confined as much as possible. Spreading it around would only create more damage. (2)

G. I don't understand. (0)

H. I don't know enough about this subject to make a choice. (0)

I. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.8 Item Number 8

A scientist's religious views will NOT make a difference to the scientific discoveries he or she makes.

Your position, basically: (Please read from A to G, and then choose ONE only from this page.)

A. Religious views do not make a difference. Scientists make discoveries based on scientific theories and experimental methods, not on religious beliefs. Religious beliefs are outside the domain of science. (2)

B. It depends on the particular religion itself, and on the strength or importance of an individual's religious views. (3)

Religious views do make a difference:

C. because religious views will determine how you judge scientific ideas. (1)

D. because sometimes religious views may affect what scientists do or what problems they choose to work on. (4)

E. I don't understand. (0)

F. I don't know enough about this subject to make a choice. (0)

G. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.9 Item Number 9

Community or government agencies should tell scientists what to investigate; otherwise scientists will investigate what is of interest only to them.

Your position, basically: (Please read from A to J, and then choose ONE only from this page.)

Community or government agencies should tell scientists what to investigate:

A. so that the scientists' work can help improve society. (2)

B. only for important public problems; otherwise scientists should decide what to investigate. (2)

C. All parties should have an equal say. Government agencies and scientists together should decide what needs to be studied, even though scientists are usually informed about society's needs. (4)

D. Scientists should mostly decide what to investigate, because they know what needs to be studied. Community or government agencies usually know little about science; their advice however, might sometimes be helpful. (2)

E. Scientists should mostly decide because they know best: which areas are ready for a break-through, which areas have the experts available, which areas have the available technology, and which areas have the greatest chance of helping society. (3)

F. Scientists should decide what to investigate, because they alone know what needs to be studied. Governments often put their own interests ahead of society's needs. (1)

G. Scientists should be free to decide what to investigate, because they must be interested in their work in order to be creative and successful. (1)

H. I don't understand. (0)

I. I don't know enough about this subject to make a choice. (0)

J. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.10 Item Number 10

Australian scientists should be held responsible for the harm that might result from their discoveries.

Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

A. Scientists should be held responsible because it's part of a scientist's job to ensure that no harm comes from a discovery. Science should cause no harm. (2)

B. Scientists should be held responsible because, if a discovery can be used for both good and bad purposes, the scientists must promote the good use and stop the bad use. (2)

C. Scientists should be held responsible because they must be aware of the effects of their experiments ahead of time. Science should cause more good than harm. (4)

D. The responsibility should be shared about equally between the scientists and society. (3)

Scientists should NOT be held responsible:

E. because it's the people who use the discoveries who are responsible. Scientists may be concerned, but they have no control over how others use their discoveries. (2)

F. because the results of scientific work can't be foreseen (we can't predict if the results will be harmful or not). It's a chance we have to take. (2)

G. because otherwise scientists would quit doing research and science would not progress. (1)

H. because once a discovery is made, others should check its effects. The scientist's job is only to make the discoveries. Science and moral questions are separate. (1)

I. I don't understand. (0)

J. I don't know enough about this subject to make a choice. (0)

K. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.11 Item Number 11

The best scientists are those who follow the steps of the scientific method.

Your position, basically: (Please read from A to H, and then choose ONE only from this page.)

A. The scientific method ensures valid, clear, logical and accurate results. Thus, most scientists will follow the steps of the scientific method. (2)

B. The scientific method should work well for most scientists; based on what we learned in school. (1)

C. The scientific method is useful in many instances. but it does not ensure results. Thus, the best scientists will also use originality and creativity. (4)

D. The best scientists are those who use any method that might get favourable results (including the method of imagination and creativity). (4)

E. Many scientific discoveries were made by accident, and not by sticking to the scientific method. (3)

F. I don't understand. (0)

G. I don't know enough about this subject to make a choice. (0)

H. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.12 Item Number 12

Within Australia there are groups of people who feel strongly in favour of or strongly against some research field. Science and technology projects are influenced by these special interest groups (such as environmentalists, religious organisations, and animal rights people).

Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

Special interest groups do have an influence:

A. because they have the power to stop some research projects and that field of science suffers. (2)

B. because they have the power to tell scientists which projects are important to do or not to do. (2)

C. because they influence public opinion and therefore the scientists. (3)

D. because they influence government policy and governments decide whether to fund a research project or not. (4)

E. because some special interest groups give money for certain research projects. (2)

F. Special interest groups try to have an influence but they don't always succeed because scientists and technologists have the final say. (2)

Special interest groups do NOT have an influence:

G. because the government decides the direction that research will take. (1)

H. because science and government decide what projects are important and they do them no matter what special interest groups say. (1)

I. I don't understand. (0)

J. I don't know enough about this subject to make a choice. (0)

K. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.13 Item Number 13

Today in Australia, there are more male scientists than female scientists. The MAIN reason for this is:

Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

A. males are stronger, faster, brighter, and better at concentrating on their studies. (1)

B. males seem to have more scientific abilities than females, who may excel in other fields. (1)

C. males are just more interested in science than females. (2)

D. the traditional stereotype held by society has been that men are smarter and dominant, while women are weaker and less logical. This prejudice has caused more men to become scientists, even though females are just as capable in science as males. (3)

E. the schools have not done enough to encourage females to take science courses. Females are just as capable in science as males. (2)

F. until recently, science was thought to be a man's vocation. (Women didn't fit television's stereotype image of scientist.) In addition, most women were expected to work in the home or take on traditional jobs. (Thus men have had more encouragement to become scientists.) But today this is changing. Science is becoming a vocation for women, and women are expected to work in science more and more. (4)

G. women have been discouraged, or not allowed, to enter the scientific field. Women are just as interested and just as capable as men; but the established scientists (who are male) tend to discourage or intimidate potential female scientists. (3)

H. There are NO reasons for having more male scientists than female scientists. Both sexes are equally capable of being good scientists, and today the opportunities are equal. (2)

I. I don't understand. (0)

J. I don't know enough about this subject to make a choice. (0)

K. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.14 Item Number 14

Science and technology can NOT help people make legal decisions; for example, deciding if a person is guilty or not guilty in a court of law.

Your position, basically: (Please read from A to G, and then choose ONE only from this page.)

Science and technology can NOT help:

A. because they have nothing to do with legal decisions, since legal decisions are based on moral values and beliefs. (1)

B. because it's wrong to base legal decisions on technology such as the lie detector. (2)

Science and technology CAN help in a number of cases:

C. by developing ways to gather evidence and by testifying about the physical facts of a case. (4)

D. by studying human behaviour and explaining the human circumstances of a case. (3)

E. I don't understand. (0)

F. I don't know enough about this subject to make a choice. (0)

G. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.15 Item Number 15

Scientific research would be better off in Australia if the research were more closely controlled by corporations (for example, companies in high-technology, communications, pharmaceuticals, forestry, mining, manufacturing).

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

Corporations should mainly control science:

A. because closer control by corporations would make science more useful and cause discoveries to be made more quickly through faster communication, better funding, and more competition. (2)

B. in order to improve the cooperation between science and technology, and thus solve problems together. (2)

C. but the public or government agencies should have a say in what science tries to achieve. (2)

Corporations should not control science:

D. because if corporations did, scientific discoveries would be restricted to those discoveries that benefit the corporation (for example, making a profit). Important scientific discoveries that benefit the public are made by unrestricted pure science. (4)

E. because if corporations did, corporations would obstruct scientists from investigating important problems which the companies wanted kept quiet; for example, pollution by the corporation. (3)

F. Science cannot be controlled by corporations. No one, not even the scientist, can control what science will discover. (1)

G. I don't understand. (0)

H. I don't know enough about this subject to make a choice. (0)

I. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.16 Item Number 16

Scientists and engineers should be the ones to decide what types of energy Australia will use in the future (for example, nuclear, hydro, solar, or coal burning) because scientists and engineers are the people who know the facts best.

Your position, basically: (Please read from A to J, and then choose ONE only from this page.)

Scientists and engineers should decide:

A. because they have the training and facts which give them a better understanding of the issue. (1)

B. because they have the knowledge and can make better decisions than government bureaucrats or private companies, both of whom have vested interests. (1)

C. because they have the training and facts which give them a better understanding; BUT the public should be involved, either informed or consulted. (2)

D. The decision should be made equally; viewpoints of scientists and engineers, other specialists, and the informed public should all be considered in decisions which affect our society. (4)

E. The government should decide because the issue is basically a political one; BUT scientists and engineers should give advice. (3)

F. The public should decide because the decision affects everyone; BUT scientists and engineers should give advice. (2)

G. The public should decide because the public serves as a check on the scientists and engineers. Scientists and engineers have idealistic and narrow views on the issue and thus pay little attention to consequences. (2)

H. I don't understand. (0)

I. I don't know enough about this subject to make a choice. (0)

J. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.17 Item Number 17

We always have to make trade-offs (compromises) between the positive and negative effects of science and technology.

Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

There are always trade-offs between benefits and negative effects:

A. because every new development has at least one negative result. If we didn't put up with the negative results, we would not progress to enjoy the benefits. (3)

B. because scientists cannot predict the long-term effects of new developments, in spite of careful planning and testing. We have to take the chance. (3)

C. because things that benefit some people will be negative for someone else. This depends on a person's viewpoint. (4)

D. because you can't get positive results without first trying a new idea and then working out its negative effects. (2)

E. but the trade-offs make no sense. (For example: Why invent labour saving devices which cause more unemployment? or Why defend a country with nuclear weapons which threaten life on earth?) (2)

There are NOT always trade-offs between benefits and negative effects:

F. because some new developments benefit us without producing negative effects. (2)

G. because negative effects can be minimised through careful planning and testing. (4)

H. because negative effects can be eliminated through careful planning and testing. Otherwise, a new development is not used. (1)

I. I don't understand. (0)

J. I don't know enough about this subject to make a choice. (0)

K. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.18 Item Number 18

The best scientists are always very open-minded, logical, unbiased and objective in their work. These personal characteristics are needed for doing the best science.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

A. The best scientists display these characteristics otherwise science will suffer. (2)

B. The best scientists display these characteristics because the more of these characteristics you have, the better you'll do at science. (2)

C. These characteristics are not enough. The best scientists also need other personal traits such as imagination, intelligence and honesty. (4)

The best scientists do NOT necessarily display these personal characteristics:

D. because the best scientists sometimes become so deeply involved, interested or trained in their field, that they can be closed-minded, biased, subjective and not always logical in their work. (2)

E. because it depends on the individual scientist. Some are always open-minded, objective, etc. in their work; while others can be come closed-minded, subjective, etc. in their work. (3)

F. The best scientists do NOT display these personal characteristics any more than the average scientist. These characteristics are NOT necessary for doing good science. (1)

G. I don't understand. (0)

H. I don't know enough about this subject to make a choice. (0)

I. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.19 Item Number 19

Politics in Australia affects Australian scientists, because scientists are very much part of Australian society (that is, scientists are not isolated from society).

Your position, basically: (Please read from A to J and then choose ONE only from this page.)

Scientists ARE affected by Australian politics:

A. because funding for science comes mainly from governments which control the way the money is spent. Scientists sometimes have to lobby for funds. (4)

B. because governments not only give money for research, they set policy regarding new developments. This policy directly affects the type of projects scientists will work on. (4)

C. because scientists are a part of society and are affected like everyone else. (3)

D. because scientists try to help society and thus they are closely tied to society. (2)

Scientists are NOT affected by Australian politics:

E. because the nature of a scientist's world prevents the scientist from becoming involved politically. (1)

F. because scientists are isolated from society; their work receives no public media attention unless they make a spectacular discovery. (1)

G. because Australia is a free country, and so scientists can work quite freely. (1)

H . I do not understand . (0)

I. I don't know enough about this subject to make a choice. (0)

J. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.20 Item Number 20

When doing science or technology, a good female scientist would carry out the job basically *in the same way* as a good male scientist.

Your position, basically: (Please read from A to L, and then choose ONE only from this page.)

There is NO difference between female and male scientists in the way they do science:

A. because all good scientists carry out the job the same way. (2)

B. because female and male scientists experience the same training. (3)

C. because overall, women and men are equally intelligent. (2)

D. because women and men are the same in terms of what is needed to be a good scientist. (3)

E. because everyone is equal, no matter what the job. (1)

F. because any differences in the way scientists do science are due to differences between individuals. Such differences have nothing to do with being male or female. (4)

G. Women would do science somewhat differently because, by nature or by upbringing, females have different viewpoints, perspectives, imagination, or characteristics (such as patience). (3)

H. Men would do science somewhat differently because men do science better. (1)

I. Women would likely do science somewhat better than men, because women must work harder in order to compete in a male dominated field such as science. (2)

J. I don't understand. (0)

K. I don't know enough about this subject to make a choice. (0)

L. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.21 Item Number 21

The success of science and technology in Australia depends on how much support the public gives to scientists, engineers and technicians. This support depends on high school students , the future public, learning how science and technology are used in Australia.

Your position, basically: (Please read from A to H, and then choose ONE only from this page.)

Yes, the more students learn about science and technology:

A. the better they will keep the country running. High school students are the future. (2)

B. the more students will become scientists, engineers and technicians, and so Australia will prosper. (2)

C. the more informed the future public will be. They will be able to form better opinions and make better contributions to how science and technology are used. (4)

D. the more the public will see that science and technology are important. The public will better understand the views of experts and will provide the needed support for science and technology. (3)

E. No, support does not depend on students learning more about science and technology. Some high school students are not interested in science subjects. (1)

F. I don't understand. (0)

G. I don't know enough about this subject to make a choice. (0)

H. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.22 Item Number 22

More technology will improve the standard of living for Australians.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

A. Yes, because technology has always improved the standard of living, and there is no reason for it to stop now. (1)

B. Yes, because the more we know, the better we can solve our problems and take care of ourselves. (2)

C. Yes, because technology creates jobs and prosperity. Technology helps life become easier, more efficient and more fun. (2)

D. Yes, but only for those who can afford to use it. More technology will cut jobs and cause more people to fall below the poverty line. (3)

E. Yes and no. More technology would make life easier, healthier and more efficient. BUT more technology would cause more pollution, unemployment and other problems. The standard of living may improve, but the quality of life may not. (4)

F. No. We are irresponsible with the technology we have now; for example, our production of weapons and using up our natural resources. (3)

G. I don't understand (0)

H. I don't know enough about this subject to make a choice. (0)

I. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.23 Item Number 23

Few scientists and technologists would choose to work on military research and development.

Your position, basically: (Please read from A to H, and then choose ONE only from this page.)

Few would do research and development for the military:

A. because many scientists and technologists would rather work in other areas which benefit human life and the environment. (2)

B. because many scientists and technologists would not sacrifice their morals and contribute to the violence of war. (2)

C. It depends on the person's values and research interests. Some scientists would find the military projects interesting and rewarding; other scientists would rather not work on projects related to war. (4)

A number of scientists and technologists choose to work for the military:

D. because most of the research money is in arms technology and military related research. The military offers large budgets. excellent equipment and more recognition for scientists. (3)

E. because they know that our country's defence is important. We need more scientists in military research and development. (1)

F. I don't understand. (0)

G. I don't know enough about this subject to make a choice. (0)

H. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.24 Item Number 24

We have to be concerned about pollution problems which are unsolvable today. Science and technology cannot necessarily fix these problems in the future.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

Science and technology can NOT fix such problems:

A. because science and technology are the reason that we have pollution problems in the first place. More science and technology will bring more pollution problems. (1)

B. because pollution problems are so bad today they are already beyond the ability of science and technology to fix them. (1)

C. because pollution problems are becoming so bad that they may soon be beyond the ability of science and technology to fix them. (2)

D. No one can predict what science and technology will be able to fix in the future. (3)

E. Science and technology alone cannot fix pollution problems. It is everyone's responsibility. The public must insist that fixing these problems is a top priority. (4)

F. Science and technology can fix such problems, because the success at solving problems in the past means science and technology will be successful in the future at fixing pollution problems. (2)

G. I don't understand. (0)

H. I don't know enough about this subject to make a choice. (0)

I. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.25 Item Number 25

Science would advance more efficiently in Australia if it were more clearly controlled by the government.

Your position, basically: (Please read from A to H, and then choose ONE only from this page.)

A. Government should control science and make it more efficient by coordinating research work and by providing the money. (3)

B. The government's control should depend upon how useful the particular scientific research will be for Australian society. Useful research should be more closely controlled and money should be provided. (4)

C. Government should NOT control science, but should give it money and leave the conduct of science up to scientists. (2)

D. Government should NOT control science but should leave the scientific research to private agencies or corporations; though government should provide the money for the scientific research. (1)

E. Government cannot make science more efficient because government is inefficient and cannot always be trusted. (2)

F. I don't understand. (0)

G. I don't know enough about this subject to make a choice. (0)

H. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.26 Item Number 26

Science and technology influence our everyday thinking because science and technology give us new words and ideas.

Your position, basically: (Please read from A to I, and then choose ONE only from this page.)

A. Yes, because the more you learn about science and technology, the more your vocabulary increases, and thus the more information you can apply to everyday problems. (2)

B. Yes, because we use the products of science and technology (for example, computers. microwaves, health care). New products add new words to our vocabulary and change the way we think about everyday things. (2)

C. Science and technology influence our everyday thinking BUT the influence is mostly from new ideas, inventions and techniques which broaden our thinking. (2)

Science and technology are the most powerful influences on our everyday lives, not because of words and ideas:

D. but because almost everything we do, and everything around us, has in some way been researched by science and technology. (4)

E. but because science and technology have changed the way we live. (3)

F. No, because our everyday thinking is mostly influenced by non-scientific things. Science and technology influence only a few of our ideas. (1)

G. I don't understand. (0)

H. I don't know enough about this subject to make a choice. (0)

I. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

Appendix 2.27 Item Number 27

Most Australian scientists are motivated to work hard. The MAIN reason behind their *personal* motivation for doing science is:

Your position, basically: (Please read from A to K, and then choose ONE only from this page.)

A. earning recognition, otherwise their work would not be accepted. (3)

B. earning money, because society pressures scientists to strive after financial rewards. (2)

C. acquiring a bit of fame, fortune and power, because scientists are like anyone else. (2)

D. satisfying their curiosity about the natural world, because they like to learn more all the time and solve mysteries of the physical and biological universe. (3)

E. solving curious problems for personal knowledge, AND discovering new ideas or inventing new things that benefit society (for example, medical cures, answers to pollution, etc.). Together these represent the main personal motivation of most scientists. (4)

F. unselfishly inventing and discovering new things for technology. (1)

G. discovering new ideas or inventing new things that benefit society (for example, medical cures, answers to pollution, etc.). (1)

H. It's not possible to generalise because the main personal motivation of scientists varies from scientist to scientist. (4)

I. I don't understand. (0)

J. I don't know enough about this subject to make a choice. (0)

K. None of these choices fits my basic viewpoint. (0) *Scores for responses are shown in parenthesis*

APPENDIX 3: THE FIT DATA FOR THE ITEMS

Appendix 3.1 Item estimates (thresholds) in input order: students on Science Scale (N = 1278; L = 10)

Item Name	Score	Max Score	Thresholds				Inft	Outft		
			1	2	3	4	MnSq	Mn Sq	Inft t	Outft t
Item 1	3000	5112	-1.06	-0.78	0.46	2.02	1.09	1.09	2.0	1.6
			.16	.13	.12	.12				
Item 2	2995	5112	-0.89	-0.70	0.68	1.05	1.10	1.09	2,5	1.9
			.13	.12	.09	.11				
Item 7	3540	5112	-0.91	-0.63	0.05	0.73	0.96	0.95	-1.0	-0.9
			.13	.12	.11	.10				
Item 10	2803	5112	-1.28	-0.70	0.90	1.40	1.10	1.10	2.3	2.0
			.14	.13	.11	.13				
Item 14	3576	5112	-0.50	-0.13	0.06	0.26	0.97	0.96	-0.8	-0.5
			.09	.11	.09	.09				
Item 16	2966	5112	-1.31	-0.15	0.58	0.73	0.94	0.95	-1.7	-1.0
			.13	.11	.10	.10				
Item 17	3239	5112	-0.89	-0.39	0.17	1.03	1.05	1.07	1.3	1.4
			.11	.10	.10	.09				
Item 22	3397	5112	-1.13	-0.32	0.21	0.59	0.92	0.90	-2.4	-2.1
			.13	.10	.09	.10				
Item 24	3277	5112	-0.91	-0.23	0.17	-.75	-.93	-.92	-2.1	-1.6
			.13	.12	.09	.09				
Item 26	2655	5112	-0.75	-0.46	0.91	1.34	1.01	1.01	0.3	0.2
			.13	.12	.12	.10				
Mean			**0.00**				**1.01**	**1.01**	**0.3**	**0.2**
SD			**0.14**				**0.07**	**0.08**	**1.9**	**1.5**

Appendix 3.2 Item estimates (thresholds) in input order: students on Scientist Scale (N = 1278; L = 8)

Item number	Score	Max Score	Threshold				Inft	Outft		
			1	2	3	4	Mn Sq	Mn Sq	Inft t	Outft t
Item 4	3229	5112	-1.25	-0.60	0.53	0.91	0.99	0.98	-0.2	-0.3
			.16	.13	.10	.09				
Item 6	3337	5112	-0.63	0.01	0.15	0.58	1.09	1.14	2.7	2.4
			.11	.09	.11	.10				
Item 8	2796	5112	-0.67	-0.18	0.59	1.34	1.07	1.08	1.8	1.6
			.11	.10	.09	.11				
Item 11	3510	5112	-0.39	-0.16	0.08	0.46	0.99	0.98	-0.3	-0.3
			.11	.10	.10	.09				
Item 13	3142	5112	-1.42	-0.70	0.52	1.28	0.95	0.95	-1.2	-1.1
			.17	.14	.09	.11				
Item 18	3058	5112	-0.69	-0.30	0.43	1.08	0.99	1.00	-0.3	-0.1
			.13	.10	.09	.11				
Item 20	3189	5112	-1.22	-0.55	0.43	1.14	0.93	0.93	-2.0	-1.6
			.16	.13	.10	.09				
Item 27	3727	5112	-0.87	-0.41	0.92	0.52	1.00	0.98	0.1	-0.3
			.13	.11	.09	.09				
Mean			**0.00**				**1.00**	**1.00**	**0.1**	**0.0**
SD			**0.15**				**0.05**	**0.07**	**1.5**	**1.3**

APPENDIX 4: MAP OF RESPONDENTS AND ITEM ESTIMATES (THRESHOLDS)

Appendix 4.1 Map of respondents and item estimates (thresholds): all students on Science Scale (N = 1278; L = 10)

2.0		**1.4**			
	x				
	xx				
		10.4	**26.4**		
	xxx				
	xxxxx				
1.0	**xxxxxxx**	**2.4**	**17.4**		
		10.3	**26.3**		
	xxxxxxxx				
	xxxxxxxxxxxxxxxxxxxxx	**2.3**	**7.4**	**16.4**	**24.4**
	xxxxxxxxx	**16.3**	**22.4**		
	xxxxxxxxxxxx	**1.3**			
	xxxxxxxxxx				
	xxxxxxxxxxxxxxxxxxx	**14.4**			
	xxxxxxxxxx	**17.3**	**22.3**	**24.3**	
0.0	**xxxxxxxxxxxxxxxxxxxx**	**7.3**	**14.3**		
	xxxxxxxxxxxx				
	xxxx	**14.2**	**16.2**		
	xxxxxx	**22.2**	**24.2**		
	xx	**17.2**			
	xxx	**14.1**	**26.2**		
	x	**7.2**			
	xx	**2.2**	**10.2**	**26.1**	
	x	**1.2**			
	x	**2.1**	**7.1**	**17.1**	**24.1**
-1.0	**x**	**1.1**			
		22.1			
		10.1			
		16.1			
-2.0					
	x				

Each X represents 8 students

Appendix 4.2 Map of respondents and item estimates (thresholds): all students on Scientists Scale (N = 1426; L = 8)

	x				
2.0					
	x				
	xx				
	xxxx	8.4			
		13.4			
	xxxxx	20.4			
1.0		18.4			
	xxxxxx	4.4			
	xxxxxxx				
	xxxxxxx				
	xxxxxxxxx	6.4	8.3		
	xxxxxxxxx	4.3	11.4	13.3	27.4
	xxxxxxxxx	18.3	20.3		
	xxxxxxxxxxx	18.3	20.3		
	xxxxxxxxxxxxxxxxxxxxx				
	xxxxxx	6.3			
0.0	xxxxxx	6.2	11.3	27.3	
	xxxxxxxxx				
	xxx	8.2	11.2		
	xxxx	18.2			
	xxxx	11.1	27.2		
	x				
	x	4.2	6.1	20.2	
	x	8.1	13.2	18.1	
	x				
-1.0		27.1			
		4.1	20.1		
		13.1			
-2.0					

Each X represents 11 students

The impact of training on rater variability

Steven Barrett
University of South Australia steven.barrett@*unisa.edu.au*

In the five years, 1993 to 1998, total Commonwealth Government spending on education fell from 4.9 to 4.4 per cent of Gross Domestic Product. Australian universities have responded to this changed funding environment though the increased use of casual teaching staff. The aim of this study was to develop, implement and evaluate a short, cost-effective training package designed to improve the rating performance of causal teaching staff. The pre and post training performance of a group of raters was measured using the Partial Credit Model, an extension of the Rasch model.

The intervention was largely unsuccessful. The study may have identified the existence of cultural barriers to the training of academic staff, both casual and tenured. This study should be repeated using a revised method and a more extensive training procedure. The participants in the proposed follow-up study should also be interviewed to identify their views about training.

rater training, rasch model, partial credit model

INTRODUCTION

During the 1980s and 1990s Commonwealth Governments of both political persuasions proudly pointed to the apparently increasing levels of government spending as proof of their commitment to education. However, in the five years from 1993 to 1998, total Commonwealth Government spending on education fell from 4.9 to 4.4 per cent of Gross Domestic Product. Moreover, Commonwealth Government expenditure per student in the higher education sector has been falling consistently since 1983. Furthermore, the rate of decline has significantly increased since the Coalition Government took office in 1996. Australian universities have responded to the changed funding environment, *inter alia*, through the increased casualisation of the teaching staff. The increased casualisation of the teaching staff has not gone unnoticed by Australian students. For example, a recent study (Barrett, 1999) found that students are concerned about the possible effects of casualisation on marker consistency. Two important results emerged from that study. First, the study identified considerable inter-rater variability and intra-rater variability. Second, the study demonstrated that the raters were constantly making four of the five common rating errors identified by Saal *et al* (1980). However, the study offered no explanation as to why the rating performance of sessional staff was significantly lower than that of tenured or contract staff.

This study aims to develop, implement and evaluate a cost-effective intervention to improve the rating performance of sessional staff. The intervention took the form of a short, that is a half-hour, training package delivered as part of the markers' meeting for a subject. The aim of the training package was to improve rater performance by reducing in the incidence of the five common rater errors identified by Saal *et al* (1980) namely, severity or leniency, the halo effect, the central tendency effect, the restriction of range effect and inter-rater reliability or agreement. The incidence of these rating errors can be measured in an Item Response Theory framework using the Partial Credit Model (Masters, 1982), which is an extension of the Simple Logistic Model (Rasch, 1960).

STANDARD SETTING JUDGES

Examination marking requires raters to make complex judgments and decisions quickly in order to meet increasingly tight end of semester deadlines. On the other side of the marking equation are students, who require raters to make consistent judgments about the minimum level of competence for each grade. Consistency in examination marking can only be expected if the raters are highly knowledgeable in the domain in which ratings are made (Jaeger, 1991). Such raters are referred to as experts. However, the increased casualisation of academe means that examinations are increasingly being marked by groups of people who vary quite markedly in their level of expertise. Consequently, raters are increasingly novices, which adversely affects the consistency of ratings.

Jaeger (1991; 4) argues that experts can be described with respect to eight criteria. First, experts excel in their own domains of knowledge. Second, experts are able to perceive large meaningful patterns in their domain of experience. Third, experts are able to perform rapidly in their domain of experience. Fourth, experts see and represent a problem in their own domain at a deeper, more principled level, than novices. Fifth, experts spend time analysing a problem qualitatively. Sixth, experts have strong self-monitoring skills. Seventh, experts are more accurate than novices at judging problem difficulty. Finally, expertise lies more in elaborated semantic memory than in a general reasoning process. Most importantly, novices provide estimates of item difficulty that are incompatible with the estimates of other raters (van der Linden, 1982). The key to ensuring consistent rater performance lies in the selection process. However, in many university departments, the field of potential raters is often limited. Hence, the challenge is to assist novices to perform like experts.

The logical solution to make novices perform like experts is training. However, this may not be as easy as it first seems as raters need to acknowledge the context in which rating occurs. In addition, marking an examination with a large number of candidates is a complex process that consists of three elements. That is, a team of raters, interacting with a set of test items, through the use of a particular standard setting process (Plake *et al* 1991). An analysis of these three elements identifies a range of factors, in addition to the expert/novice dichotomy, which may affect intra-rater consistency.

The first source of intra-rater inconsistency is a range of factors that are related to the raters themselves. This is not surprising as any team of raters will differ with respect to experience, specialties and professional skill. Individual raters may also have idiosyncratic perceptions about the knowledge or skills that are required to demonstrate the minimum level of competence for a particular test item. This is more likely to be a cause of concern if the examination contains items that test a broad range of skills or knowledge. Furthermore, inconsistencies in rater performance may be exacerbated by fatigue during the rating process (Plake *et al*, 1991).

A second set of factors that may lead to intra-rater inconsistencies are related to the items and the examination. The perceptions of raters about the quality of items or the appropriateness of an examination may lead to more inconsistent ratings. Plake *et al* (1991) cites the example of raters who disagreed about the validity of an examination for certification purposes. Raters who felt that the examination was not valid were less conscientious and more prone to lapses of concentration, thereby accentuating the fatigue effect. Plake *et al* (1991) also argued that the rater factors and the examination factors may interact with each other to produce a third source of inconsistencies. For example, novice raters may be less consistent when marking long examinations that contain complex and demanding items.

Finally, there are a number of factors relating to the rating process itself. For example, the absence of a marking guide may be a source of inconsistency when raters are confronted with unfamiliar content. Furthermore, rater inconsistency may result if the group of raters is unable to meet and discuss the rating process beforehand.

Plake *et al* (1991) argue that there are five strategies that can be used to improve rater inconsistencies.

> **Periodic retraining** The rating process is periodically interrupted to conduct additional group discussion. These discussions ensure that the raters maintain consistent definitions of the minimum competent candidates.
>
> **Estimations of minimally competent test performance** This involves providing raters with empirical data relating to the performance of previous candidates on similar or identical test items. Such information can range from simply providing raters with information about the proportion of previous candidates who passed a particular item in the past, to providing raters with estimates of the person statistics from analysing previous examinations using Item Response Theory.
>
> **Empirical data on item performance** Raters can be provided with data relating to the difficulty of individual items in an examination. Again this information can range from pass rates of test items to estimates of item parameters using Item Response Theory.
>
> **Providing descriptive data relating to the performance of raters** Raters can be provided with information about the distribution of marks for the entire rating team. This requires raters to provide information to the entire group on two occasions. After receiving the first batch of information, raters should review all of their previous ratings in light of the judgments made by the other raters. Raters could use the information shared the second time for a variety of purposes, such as to review further their rating or as a basis for attaining a group consensus.

Training is a necessary condition if rater inconsistencies are to be minimised, if not eliminated. Mills, Melican and Ahluwalia (1991) argue that training of raters should achieve four important outcomes. First, training provides a context within which the rating process occurs. Second, training defines the tasks to be performed by the raters. Third, training minimises the effects of variables other than item difficulty from the rating process. Fourth, training develops a common definition of the minimally competent candidate. Furthermore, there are three measurable criteria that can be used to determine whether a rater is well trained (Reid, 1991). First, ratings should be stable throughout the rating process. Second, ratings should reflect the relative difficulties of the test items. Third, ratings should reflect realistic expectations of the expected performance of the candidates. However, the big question remains, how should raters be trained? Hambleton and Powell (1983) argue that this is a difficult question to answer due to the poor documentation of training procedures in most of the reports of standard setting studies. Nevertheless, this brief review of the literature provided a framework within which the intervention that was at the centre of the present study was developed.

THE PARTIAL CREDIT MODEL

The examination results that form the basis of this study were analysed using the computer program Conquest (Adams and Khoo, 1993). This program fits item response and latent regression models to data obtained from both dichotomously scored and polychotomously scored tests (Wu, Adams and Wilson, 1998). The data were analysed using the Partial Credit Model (Masters, 1982), which is an extension of the Simple Logistic Model (Rasch, 1960). The Simple Logistic Model is only appropriate where items are dichotomously scored, such as in true/false or multiple-choice tests. Whereas the Partial Credit Model facilitates the analysis of cognitive or attitudinal items that have two or more levels of response. The levels of response have to be ordered, but they do not have to be on a specified scale. Hence, the Partial Credit Model is ideal for analysing the effects of student ability and item difficulty on the performance of students answering extended response type questions. Moreover, the Partial Credit Model converts the ordered category scores to interval scaled scores.

Rasch (1960) developed a latent trait model for dichotomously scored items. All statistical models that are used to operationalise Item Response Theory specify a relationship between the observed performance of examinees on a test and unobservable or latent traits that are assumed to underlie the observed performance. This relationship is the item characteristic curve (Hambleton 1989). When test data fit the Rasch model, the requirements that underlie Item Response Theory have been met. The Rasch model produces item-free estimates of student ability or performance and sample-free or person-free estimates of the item parameters. That is, the Rasch model is independent of both the items on the test and the sample of people to whom the test is administered. Moreover, the Rasch model can be used to equate readily the performance of different students answering different items on a test, which replaces the concept of parallel test forms that characterises classical test theory.

The Rasch model (Rasch, 1960) estimates the probability of an examinee gaining a correct answer to a dichotomously scored item as an exponential function of the difference between the ability of a person and item difficulty. The Simple Logistic Model can be expressed as;

$$\pi_{1ni} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$$

Where: π_{1ni} is the probability for person *n* of success on item *i*,

β_n is the ability of person *n*,

δ_i is the difficulty of item *i*, and

$\pi_{0ni} = 1 - \pi_{1ni}$ is the probability of an incorrect answer on item *i*.

This is the only latent trait model for dichotomously scored responses for which the number of successes, r_n, is a sufficient statistic for the person parameter β_n (Masters, 1982; 152).

The general applicability of the Simple Logistic Model (Rasch, 1960) is greatly reduced as not all test data are dichotomously scored. Masters (1982) argues that there are four other observation formats that record ordered levels of responses.

Repeated trials The data are obtained from a fixed number of independent attempts at each item on a test.

Counts There is on upper limit to the number of independent successes or failures a person can make on an item.

Rating scales Respondents are presented with a fixed set of ordered response alternatives that are used with every item.

Partial credit Data are obtained from a test that required the prior identification of several ordered levels of performance on each item and where partial credit is awarded for partial success on items.

The Partial Credit Model developed by Masters (1982) is an extension of the Simple Logistic Model, which overcomes this substantial shortcoming. The model was developed by estimating parameters for the difficulties associated with a series of performance levels within each item. Masters (1982) argues that the difficulty of the k^{th} level in an item governs the probability of responding in category k rather than in category k - 1. The probability of person n of completing the k^{th} level is specified by Masters (1982; 158) as:

$$\pi_{xni} = \frac{\exp \sum (\beta_n - \delta_i)}{\sum \exp \sum (\beta_n - \delta_i)} \qquad x = 0,1, \ldots m,$$

where for notational convenience $\sum (\beta_n - \delta_i) = 0$.

The model estimates the probability of a person n scoring x on the mi performance level of item i as a function of the person ability on the variable being measured and the difficulties of the mi levels in item i. The observation x is a *count* of the successfully *completed* item levels, while only the difficulties of these completed levels appear in the numerator of the model. The model provides estimates of person ability β_n and level difficulty δ_i.

METHODS

Subjects

The Division of Business and Enterprise at the University of South Australia requires all undergraduate students to take a "core" of eight subjects. One of these eight subjects is Economic Environment, a principles of macroeconomics subject. Approximately 1,200 students commenced this subject in Semester 1, 1999, of whom 810 sat the final examination. Of these students, 100 students went on to complete Business Economics, a principles of microeconomics subject. Despite the obvious difference in content between these two subjects, the students were taught and assessed by the same group of staff. Hence, this study was a comparison of the rating performance of those staff members who marked both the Economic Environment Semester 1, 1999 and Business Economics, Semester 2, 1999 final examinations. Student performance on the Semester 1 examination was assessed by eight markers, three of whom were employed to mark the Semester 2 examination. The Semester 2 markers are an interesting group of three people as they include the subject convener and two sessional tutors.

The script books in this study were randomly allocated to raters, who marked all items on the paper. No crossover occurs when raters mark items that are not marked by other raters or when raters only mark the work of their own students. Whereas crossover between items, students and raters is maximised when raters mark a random sample of all papers and mark all items. Maximised crossover ensures that the Partial Credit Model fully separates the rater, student and item effects (Barrett, 1999).

The intervention

The aim of this study is to develop, implement and evaluate a short training package to improve the rating performance of sessional staff prior to the marking of the Semester 2 examination. Markers meetings in the Division of Business and Enterprise tend to be rather brief and informal affairs. The main items under consideration are the distribution of script books, a brief discussion about the marking guide and the establishment of deadlines. The intervention that was evaluated in this study was a 30 minute training session that was conducted as part of the markers' meeting for Business Economics. The training package comprised three parts, which addressed four of the five strategies for improving intra-judge consistency as reported by Mills, Melican and Ahluwalia (1991).

The first component of the training package was a presentation by the author to the raters about the nature of the five common rater errors (Saal *et al*, 1980). The aim was to sensitise the raters to the types of errors they were committing.

The second component was a discussion of the performance of the people who marked the 1998 Business Economics and the Semester 1, 1999 Economic Environment examinations. This discussion was based on the results of the Partial Credit Model analysis of the marking of these examinations, and introduced the three raters to the concept that student performance is the outcome of complex interactions between student ability, rater performance and item difficulty, which could be separated from each other using Item Response Theory. This phase of the training package concluded with a discussion of the performance of each rater during the Semester 1 examination for Economic Environment.

The third component was a new style of marking guide that was developed in conjunction with the subject convener. Previous marking guides tended to focus on content with marks being awarded for particular points. Such marking guides did not reward answers that were qualitatively better than others. They also penalised candidates who took a different approach to answering questions. Consequently, the subject convener developed a marking guide that outlined the minimum level of achievement for the grades of pass, credit and distinction.

RESULTS

The aim of this study was to evaluate the effectiveness of a training package designed to reduce the incidence of the five common rater errors identified by Saal *et al* (1980), namely (a) leniency or severity, (b) halo effect for a person or an item, (c) central tendency effect, (d) restriction of range and (e) inter-rater reliability or agreement. Figure 1, summarises the rating performances of the eight raters who marked the Semester 1 examination for Economic Environment. The figure clearly shows two groups of raters. Raters 2,5,7 and 8 were more severe than Raters 1,3,4, and 6. The severe raters tended to be more experienced university teachers. This group of raters included the convener of Economic Environment, the convener of Business Economics and two highly experienced sessional staff who have previously held academic appointments. Conversely, raters 1, 3 and 4 were relatively inexperienced sessional staff who had only recently completed their honours degrees. Paradoxically, Rater 6 was a long serving tenured member of staff who had previously been the convener of both Economic Environment and Business Economics.

The item estimates shown in Figure 1 indicate that the three essay questions on the Economic Environment paper were all approximately of the same level of difficulty. This is quite unusual, as examinations tend to contain items that vary in difficulty. The absence of variation in item difficulty is reflected in the estimates of the rater*item interaction. However, the disturbing point shown in Figure 1 is that these items are too difficult for the majority of students. Figure 2, summarises the rating performances of the three raters who marked the Semester 2 examination for Business Economics. The variations in the item estimates shown in Figure 2 are more typical of an essay style examination. Furthermore, the figure also shows that the difficulty of these four items is more appropriate for this group of students. The vertical scale of Figure 1 and 2 is an interval scale, the units of which are logits. Some parameters could not be shown on these figures. In Figure 1 each “x” represents 17.6 students and in Figure 2 each “x” represents 2.5 students.

The rater estimates reported in Table 1 and shown in the rater column of Figure 1, indicate that on average Rater 2 was a harder marker than both sessional markers before the training was undertaken. However, the rater estimates reported in Table 2 shows that there is some variation in the severity of rating for individual items. Both sessional staff (Raters 1 and 3) have marked Item 1 harder than Rater 2, while Rater 1 was the hardest marker for Item 3. The post training estimates for rater severity reported in Table 1 shows that on average the sessional markers were more severe that the subject convener. This paradox may have been the result of the two sessional raters experiencing some performance anxiety.

Evidence of the extent of inter-rater reliability or agreement between the raters is best obtained from inspecting the rater*item columns of Figures 1 and 2. These figures are produced from the tables of rater*item estimates that are provided by Conquest. This type of error would be absent if each rater has correctly estimated the difficulty of each item. In Figure 1, the range of item difficulties for the Economic Environment examination is only 0.018 logits, but the range of rater*item estimates is 0.124, which is clearly greater. This increase is largely due to Rater 8 marking Item 1 as if it were much harder, while marking Items 2 and 3 as if they were much easier items. Rater 5 also marked Item 2 as if it were much more difficult. Figure 1 therefore suggests that with the exception of Rater 8, and to some extent Rater 5, there was strong inter-rater reliability or agreement, that is consistency, between the ratings of the group of people who marked the Economic Environment examination.

```
                           Terms in the Model Statement
Students              raters               items                rater*item
-----------------------------------------------------------------------------------------
 0.5            |                 |                 |                      |
                |                 |                 |                      |
               X|                 |                 |                      |
               X|2                |                 |                      |
               X|5                |                 |                      |
               X|7                |                 |                      |
               X|8                |                 |                      |
               X|                 |                 |                      |
               X|                 |                 |                      |
               X|                 |                 |8.1 5.2 6.2 1.3       |
               X|                 |                 |3.1 7.1 2.2 4.2       |
   0          XX|                 |1 2 3            |1.1 6.1 1.2 3.2       |
              XX|                 |                 |2.1 4.1 5.1 7.2       |
          XXXXXX|                 |                 |8.2 8.3               |
      XXXXXXXXXX|                 |                 |                      |
         XXXXXXX|                 |                 |                      |
         XXXXXXX|                 |                 |                      |
              XX|1 3 4 6          |                 |                      |
               X|                 |                 |                      |
                |                 |                 |                      |
                |                 |                 |                      |
                |                 |                 |                      |
 -0.5           |                 |                 |                      |
-----------------------------------------------------------------------------------------
```

Figure 1: Economic Environment Semester 1, 1999, Map of Latent Distributions and Response Model Parameter Estimates.

```
                          Terms in the Model Statement
students             raters                items                rater*item
 0.5            |                 |                 |                      |
                |                 |                 |                      |
                |                 |                 |                      |
               X|                 |                 |                      |
                |                 |                 |                      |
               X|                 |                 |                      |
               X|                 |                 |2.4                   |
               X|                 |                 |                      |
              XX|                 |                 |3.1 3.3               |
             XXX|                 |4                |                      |
          XXXXXX|                 |                 |1.2 2.2               |
      XXXXXXXXXX|1 3              |1                |                      |
           XXXXX|                 |3                |1.1 1.3               |
   0       XXXXX|                 |                 |                      |
              XX|                 |                 |                      |
               X|                 |                 |                      |
                |2                |                 |1.4 3.4               |
                |                 |                 |2.3                   |
                |                 |2                |2.1 3.2               |
                |                 |                 |                      |
                |                 |                 |                      |
                |                 |                 |                      |
                |                 |                 |                      |
                |                 |                 |                      |
                |                 |                 |                      |
                |                 |                 |                      |
 -0.5           |                 |                 |                      |
```

Figure 2: Business Economics Semester 2, 1999, Map of Latent Distributions and Response Model Parameter Estimates.

Table 1: Estimates of Rater Parameters

	Economic Environment Examination Semester 1, 1999			**Business Economics Examination Semester 2, 1991**		
	Rater estimates	**MNSQ**	**T**	**Rater estimates**	**MNSQ**	**T**
Rater 1	-0.279	2.20	8.3	+0.053	1.20	1.1
Rater 2	+0.320	2.17	5.8	-0.126	2.74	6.1
Rater 3	-0.275	2.53	10.3	+0.073	*	*

An asterisk next to a parameter estimate indicates that it is constrained.

Table 2: Estimates of Rater by Item Parameters

		Economic Environment Examination Semester 1, 1999			**Business Economics Semester 2, 1991**		
	Item	**Rater estimates**	**MNSQ**	**T**	**Rater estimates**	**MNSQ**	**T**
Rater 1	Item 1	-0.018	0.57	-4.7	0.028	0.56	-3.1
	Item 2	-0.006	0.95	-0.4	0.094	0.74	-1.5
	Item 3	0.024	*	*	0.017	1.18	1.0
	Item 4				-0.139	*	*
Rater 2	Item 1	-0.029	0.88	-0.7	-0.200	1.33	1.5
	Item 2	0.036	1.42	2.5	0.112	1.45	1.7
	Item 3	-0.007	*	*	-0.180	2.43	4.8
	Item 4				0.267	*	*
Rater 3	Item 1	0.031	0.75	-2.5	0.171	*	*
	Item 2	-0.024	1.12	1.1	-0.206	*	*
	Item 3	-0.007	*	*	0.163	*	*
	Item 4				-0.128	*	*

An asterisk next to a parameter estimate indicates that it is constrained.

Figure 2 demonstrates that the previously existing broad inter-rater agreement or reliability is largely absent after the training. The increased variation in the difficulty of the items on the Business Economics examination appear to have been translated into a greater dispersion of the rater*item estimates. The variation in item difficulty is 0.323 logits, whereas the range of the rater*item estimates is 0.473. This reduction in inter-rater reliability or agreement stems from the inability of the raters to estimate accurately the difficulty of these items. For example, Rater 1 correctly estimated the difficulty of Items 1 and 3 on the Economic Environment examination and marked them accordingly, but she reversed the order of the hardest and easiest items on the Business Economics examination. That is, she marked Item 4 as if it were the easiest (not the hardest) item and marked Item 2 as if it was the hardest (not the easiest item). These observations suggest that the items on an examination paper should all be of the same level of difficulty in order to achieve the highest possible level of inter-rater reliability or agreement when the rating team includes people who are not experts.

CONCLUSIONS

Since the mid-1990s, the real level of Commonwealth funding for university places has been falling. The university sector has responded to these cuts in a variety of ways. Increased employment of inexperienced sessional staff has been a fairly universal response by universities. Students are concerned that increased casualisation has led to a reduction in marker consistency. The aim of this study was to develop, implement and evaluate a cost-effective training package designed to reduce the incidence of several common rater errors. The study identified the widespread presence of only two rater errors, a marked variation in rater severity or leniency and a lack of inter-rater reliability or agreement. The lack of agreement between the raters lends support to student concerns about the lack of consistency between markers. Furthermore, it would appear that this training package did little, if anything, to improve the performance of the sessional staff raters. The only area of improvement observed was that sessional staff members were rating more severely than the subject convener after the training, rather than being more lenient as was the case before the intervention. The apparent lack of success of this study may be explained in terms of the attitudes of academics to training and shortcomings with the study design.

The issue of training academics to make careful and consistent ratings has received very little attention in the past. Indeed, markers in universities do not expect such training to be incorporated into a rating exercise, such as marking a final examination. Hence, there are considerable cultural barriers to overcome before academics are likely to accept the need for training as part of a standard setting exercise. Second, it is not clearly understood how the training of academics should be undertaken as the extensive literature on the topic relates primarily to school teachers. Nevertheless, it is clear that the time of 30 minutes that was allowed for this training package was inadequate. Furthermore, the training may have produced some performance anxiety on the part of the subjects, which might explain the paradoxical increase in severity. Third, suggestions for activities to be incorporated into a training program for academics would include a detailed scoring breakdown for each item, systematic cross checking of rater performances and detailed discussions between raters of their expectations for each item. However, it is not possible to undertake such activities when there are large numbers of students, large numbers of raters and tight deadlines. The training of raters is important. Unfortunately, no money is being spent by Australian universities on reviewing the critical process of evaluating marking procedures. Clearly more than training a small number of disparate groups of raters is required. The culture of the higher education sector needs to be changed.

A second reason why this study did not achieve its goals may be due to shortcomings in its design. The study analysed the performance of both a cohort of students and a small group of raters over an academic year. It was decided that this study design would eliminate the confounding effects that are generated from studying two different groups of raters or students. Moreover, the small number of items marked by each rater may have provided inaccurate estimates of the item and rater parameters. The study could have compared the performance of the large group of raters who marked the Economic Environment examinations at the end of Semester 1 and then at the end of Semester 2. In which case the sample sizes would be about 850 and 400 respectively. Furthermore, the number of raters being evaluated would rise to eight, which would reduce the number of parameter estimates that were constrained and hence, the amount of missing data would be greatly reduced. Clearly in fairness to students much more work should be undertaken to examine the processes used for marking in universities and to improve marker performance by rigorous and informed training of markers.

REFERENCES

Adams, R.J. and Khoo S-T. (1993) *Conquest: The Interactive Test Analysis System,* ACER Press, Hawthorn.

Barrett, S.R.F. (1999) Question choice and Marker Variability: Insights From Item Response Theory, *Unfolding Landscapes in Engineering Education*, Proceedings 11th Australasian Conference on Engineering Education, pp. 240-245, University of South Australia, September 1999.

Engelhard, G.Jr (1994) Examining Rater Error in the Assessment of Written Composition With a Many-Faceted Rasch Model, *Journal of Educational Measurement*, 31(2), 179-196.

Engelhard, G.Jr and Stone, G.E. (1998) Evaluating the Quality of Ratings Obtained From Standard-Setting Judges, *Educational and Psychological Measurement*, 58(2), 179-196.

Hambleton, R.K. (1989) Principles of Selected Applications of Item Response Theory, in R. Linn, (ed.) *Educational Measurement, 3rd ed.*, MacMillan, New York, 147-200.

Jaeger, R.M. (1991) Selection of Judges for Standard-Setting, *Educational Measurement: Issues and Practice*, 10(2), 3-10.

Keeves, J.P. and Alagumalai, S. (1999) New Approaches to Research, in G.N. Masters and J.P. Keeves, *Advances in Educational Measurement, Research and Assessment*, 23-42, Pergamon, Amsterdam.

Masters, G.N. (1982) A Rasch Model for Partial Credit Scoring, *Psychometrika*, 47, 149-174.

Mills, C.N., Melican, G.J. and Ahluwalia, N.T. (1991) Defining Minimal Competence, *Educational Measurement: Issues and Practice,* 10(2), 7-14.

Plake, B.S., Melican, G.J. and Mills, C.N. (1991) Factors Influencing Intrajudge Consistency During Standard-Setting, *Educational Measurement: Issues and Practice*, 10(2), 15-26.

Rasch, G. (1960) *Probabilistic Models for Some Intelligence and Attainment Tests,* University of Chicago Press, Chicago.

Reid, J.B. (1991) Training Judges to Generate Standard-Setting Data, *Educational Measurement: Issues and Practice*, 10(2), 11-14.

Saal, F.E., Downey, R.G. and Lahey, M.A (1980) Rating the Ratings: Assessing the Psychometric Quality of Rating Data, *Psychological Bulletin*, 88(2), 413-428.

van der Linden, W.J. (1982) A Latent Trait Method for Determining Intrajudge Inconsistency in the Angoff and Nedelsky Techniques of Standard Setting, *Journal of Educational Measurement*, 19(4), 295-308.

Wu, M.L., Adams, R.J. and Wilson, M.R. (1998) *ACER Conquest: Generalised Item Response Modelling Software*, ACER Press, Hawthorn.

School absence and student background factors: A multilevel analysis

Sheldon Rothman
Massachusetts Department of Education *srothman@doe.mass.edu*

As part of regular collections, South Australian government schools provide data on students, including individual student absences during one full term (usually 10 weeks). These data were analysed to understand how student absence is affected by student background and school contexts. A multilevel statistical model of student absence was developed using data collected in 1997, and repeated for 1999. This paper presents the findings for students in primary schools, showing that absence rates for indigenous students, while higher than the rates for non-indigenous students, are affected by school factors such as the concentration of indigenous students in the school and school socioeconomic status.

student attendance, student absence, multilevel models, socioeconomic status, indigenous students

Introduction

Regular attendance is an important factor in school success. Students who are chronic non-attenders receive fewer hours of instruction; they often leave education early and are more likely to become long term unemployed, homeless, caught in the poverty trap, dependent on welfare, and involved in the justice system (House of Representatives 1996, p. 3). High rates of student absenteeism are believed to affect regular attenders as well, because teachers must accommodate non-attenders in the same class. It has been suggested that chronic absenteeism is not a cause of academic failure and departure from formal education, but rather one of many symptoms of alienation from school. Chronic absenteeism, truancy and academic failure may be evidence of a dysfunctional relationship between student and school, suggesting that schools need to be more student-centred and supportive of students with different needs. This argument is supported by research that highlights significant associations between student background factors, poor attendance, and early school leaving (Altenbaugh, et al. 1995; Bryk & Thum 1989; Fernandez & Velez 1989).

Previous research has concentrated on students who are "chronic" or "persistent" non-attenders, examining family, academic and social background factors related to the student. Other research has concentrated on schools with high absence rates, examining student composition, school "climate" and other organisational factors associated with these rates. What has been missing is a combination of these two approaches, because the computational technology has not been available.

A European perspective on student absences was provided in a study of absenteeism in 36 high schools in four Dutch cities. Bos, Ruijters and Visscher (1992) examined aspects of absences for individual classes over three school days, a Monday, Wednesday and Friday, covering a total of 8,990 lessons. They differentiated between truancy (disallowed absence, one "without a reason that is considered valid by the school") and allowed absences (one "regarded as valid by the school"). They found variation by school in the determination of a truancy, but calculated overall absence rates of 9.1 per cent, comprising a 4.4 per cent truancy rate and a 4.7 per cent allowed absence rate. Truancy rates were lower in pre-university tracks than vocational education tracks, highest on Fridays, and tended to be higher later in the school day. Whole-day truancy occurred more frequently on Mondays. The proportion of “non-Dutch” students in the school accounted for 42 per cent of the variance in school truancy rate. The authors used schools’ administrative data to get a snapshot of truancy, reporting valuable information about truancy and absenteeism in general.

DeJung and Duckworth (1986) reported on a study of absences in two cities in the western United States. Examining data from six high schools on class absences rather than whole-day absences, they calculated absence rates of 15 per cent for the larger of the two districts, and 10 per cent for the smaller. When using whole-day absences only, rates were 4.4 per cent for the larger district and 2.8 per cent for the smaller. The researchers also asked students why they were absent from individual class periods. Of the 1,200 students in the sample, 20 per cent of students stated that they had “other things to do,” rather than attend school for a day; illness and personal problems accounted for less than 10 per cent of absences. Students with very high absence rates identified parties, drugs and a general dislike of school for most of their absences.

Throughout the 1970s, American high school principals consistently identified poor attendance as the major problem facing secondary school administrators. But rather than define poor attendance, studies concentrated on examining factors associated with it. Wright (1978) analysed secondary school-level data in Virginia, surveying schools on their attendance rates and aspects of the curriculum, organisation and staff. He found statistically significant differences by location: urban schools had the lowest attendance rates, then suburban schools; schools in other areas had the highest attendance rates. Within these geographical groupings, different factors were related to attendance rates, including subject offerings (electives), work programs for school credit, and age of the teaching staff.

Reid (1982), using data from an urban comprehensive school in a disadvantaged area of Wales, examined social background factors and self-concept in “persistent” absentees, whom he defined as students with absence rates of 65 per cent of every school term, and control groups of matched students, who were “good attenders, usually making 100 per cent attendance during an average term.” He found differences in family structure, father’s occupation, mother’s employment and occupation, and eligibility for free school meals. Of the three groups in the study, persistent absentees also scored lowest on the Brookover scale of academic self-concept, and lowest on the Coopersmith scale of self-esteem, with no differences between male and female absentees.

Two high schools in Ontario, Canada, contributed data on 54 students to a study to determine the influence of personal, family and school factors on absenteeism. Corville-Smith, Ryan, Adams and Dalicandro (1998) used discriminant analysis to identify which factors could identify truants. Perceptions of school and parental discipline and control were found to be significant factors, as were students’ perceptions of family conflict, academic self-concept and social competence in class. Unfortunately, their sample was severely restricted by selection bias: only 27 of a possible 295 volunteered to participate, and more than two-thirds were female.

Some researchers have attempted to examine the influence of attendance on academic achievement. In 1923, Odell (1923) reported small, non-significant correlations between attendance and either academic achievement or intellectual development, but significant correlations between attendance and grades awarded by teachers for class work. Finch and Nemzek (1935) reported that school grades were related to student attendance for the 1934 graduating class at one high school in Minneapolis, Minnesota. Kersting (1967) compared attendance records for the 100 highest achieving and 100 lowest achieving students in the junior high school where he was teaching. Comparing these extreme groups, he found significant differences in attendance. These studies show that while there may be a relationship between attendance and achievement, it is very poor attenders whose achievement is low, but no threshold absence rate is defined.

Research on student attendance points to some groups of students whose attendance record, as a group, is relatively poor, such as the “non-Dutch” students reported by Bos, Ruijters and Visscher (1992). For most collections of student attendance data in Australia, however, such information has not been available. Most education departments limit their annual end-of-year collections to absences at the school level, with no differentiation by any student factors. In 1997, South Australia began an annual collection of data on student absences during one ten-week term. This paper provides an analysis of these data, supplementing a summary report provided to schools and education department officials (Rothman 1999).

Data

In South Australia, government schools have the capacity to monitor student attendance electronically using computers and software. This software, called EDSAS, allows schools to record the date, type and reason for each student non-attendance.[1] Four types of non-attendance can be recorded: whole-day, morning, afternoon, and late. Sixteen reasons can be recorded, nine of which count as absences. The others, such as sport excursions and work experience, are acceptable reasons for which the student is considered present. This information can then be matched with student information to provide a rich picture of attendance and non-attendance patterns. Available student information, as provided by the school as unit records during the midyear census, includes grade (year level), date of birth, sex, indigenous status, socioeconomic status, and special need.[2]

The data in this paper were collected from schools that use EDSAS to monitor student attendance. For this paper, only whole-day absences for full-time students were used. When absence rates are discussed, the sample was limited to those students who were enrolled at one school for the entire term. The number of students and schools included each year are contained in Table 1. Comparative enrolment data are from the midyear census, conducted each year on the first Friday on or after 1 August and reported in the National Schools Statistics Collection (Australian Bureau of Statistics 1998).

In 1997 and 1998, Term 2 began after the Anzac Day holiday and was ten weeks long. There were two Monday holidays—Adelaide Cup Day (Week 4) and Queen’s Birthday (Week 7)—bringing the total number of school days to 48. Term 2 started one week earlier in 1999; with Monday holidays for Anzac Day (Week 2), Adelaide Cup Day (Week 5) and Queen’s Birthday (Week 9), there were 52 school days.

The data contained in this paper are from the 1997 and 1999 collections of individual student absences. To ensure consistency for the analysis, the files were trimmed to include only primary level full-time students who attended a single school for the entire term, resulting in 67,732 students in 304 schools in 1997, and 84,820 students in 411 schools in 1999.

Because the data are based on administrative collections, there are limits to the student-level and school-level variables that are included. Student-level variables include sex (SEX, male=0, female=1), indigenous background (ABOR, indigenous=1), SES (CARD, low SES=1), and grade level (Reception to Year 7). School-level variables include location (LOCATION, metropolitan=0, country=1), size (SIZE), per cent indigenous students (PCTABOR), per cent low SES students (PCTCARD), and per cent female students (PCTFEM). Other school indicators (Commonwealth Literacy Program or Country Areas Program school) were eliminated because of their similarity to other school-level variables. Grade level was eliminated because there was little variation by grade level across schools. Frequencies and summary statistics for the files are listed in Table 1.

Table 1. Summary statistics of variables, 1997 and 1999

	1997 (48 days)			1999 (52 days)		
Student-level variables	*n*	*Per cent of sample*	*Absences per student*	*n*	*Per cent of sample*	*Absences per student*
Sample total	**67,732**	**100.0**	**2.9**	**84,820**	**100.0**	**3.3**
Grade level						
Reception	6,924	10.2	3.5	9,065	10.7	3.8
Year 1	8,260	12.2	3.1	10,485	12.4	3.4
Year 2	8,460	12.5	2.8	10,784	12.7	3.2
Year 3	8,853	13.1	2.7	10,672	12.6	3.0
Year 4	8,628	12.7	2.8	10,994	13.0	3.0
Year 5	8,814	13.0	2.7	10,990	13.0	3.0
Year 6	8,981	13.3	2.9	10,886	12.8	3.3
Year 7	8,812	13.0	3.2	10,944	12.9	3.5
Sex						
Male	34,981	51.6	2.9	43,821	51.7	3.2
Female	32,751	48.4	3.0	40,999	48.3	3.3
Indigenous background						
Non-indigenous	65,755	97.1	2.8	82,320	97.1	3.1
Indigenous	1,977	2.9	7.4	2,500	2.9	7.3
Socioeconomic background						
Middle/upper SES	41,470	61.2	2.6	54,421	64.2	2.9
Lower SES	26,262	38.8	3.4	30,399	35.8	4.0
Location [a]						
Country	25,800	38.1	3.3	28,970	34.2	3.5
Metropolitan	41,932	61.9	2.7	55,850	65.8	3.1
School-level variables	*Mean*	*Minimum*	*Maximum*	*Mean*	*Minimum*	*Maximum*
Absences per student [b]	3.0	0.6	11.1	3.3	0.2	11.7
Size (Students R-7)	222.8	12	757	206.4	11	812
Per cent female	47.9	29.6	66.7	48.0	30.0	70.8
Per cent indigenous	3.2	0.0	84.0	3.6	0.0	100.0
Per cent low SES	40.8	0.3	88.5	37.3	4.2	90.5
Number of country schools		161	53.0%		202	49.1%
Number of metropolitan schools		143	47.0%		209	50.9%

[a] Location was used as a school-level variable only.

[b] Absences per student is an unweighted measure. For the weighted average, see the rate for student-level variables.

Methodology

It was assumed that individual student absences were influenced by student characteristics, such as sex, indigenous background and low socioeconomic status, in the context of the school the student attended. The relationship among the variables can be denoted as

$$\text{TOTABS} = \beta_{0j} + \beta_{1j}(\text{SEX}) + \beta_{2j}(\text{ABOR}) + \beta_{3j}(\text{CARD}) + \varepsilon_{ij}$$

at the student level, and

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{LOCATION}) + \gamma_{02}(\text{PCTABOR}) + \gamma_{03}(\text{PCTCARD}) + \gamma_{04}(\text{SIZE}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{LOCATION}) + \gamma_{12}(\text{PCTABOR}) + \gamma_{13}(\text{PCTCARD}) + \gamma_{14}(\text{SIZE}) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{LOCATION}) + \gamma_{22}(\text{PCTABOR}) + \gamma_{23}(\text{PCTCARD}) + \gamma_{24}(\text{SIZE}) + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}(\text{LOCATION}) + \gamma_{32}(\text{PCTABOR}) + \gamma_{33}(\text{PCTCARD}) + \gamma_{34}(\text{SIZE}) + u_{3j}$$

at the school level. Preliminary analysis showed that because the percentage of students by sex is generally within a narrow range for primary schools, the variable PCTFEM could be excluded from the model. Each of the student-level variables was grand-centred around the mean, so that the intercept term would represent the estimated mean number of days absent for schools, assuming that each school enrolled students with all the same student-level characteristics.

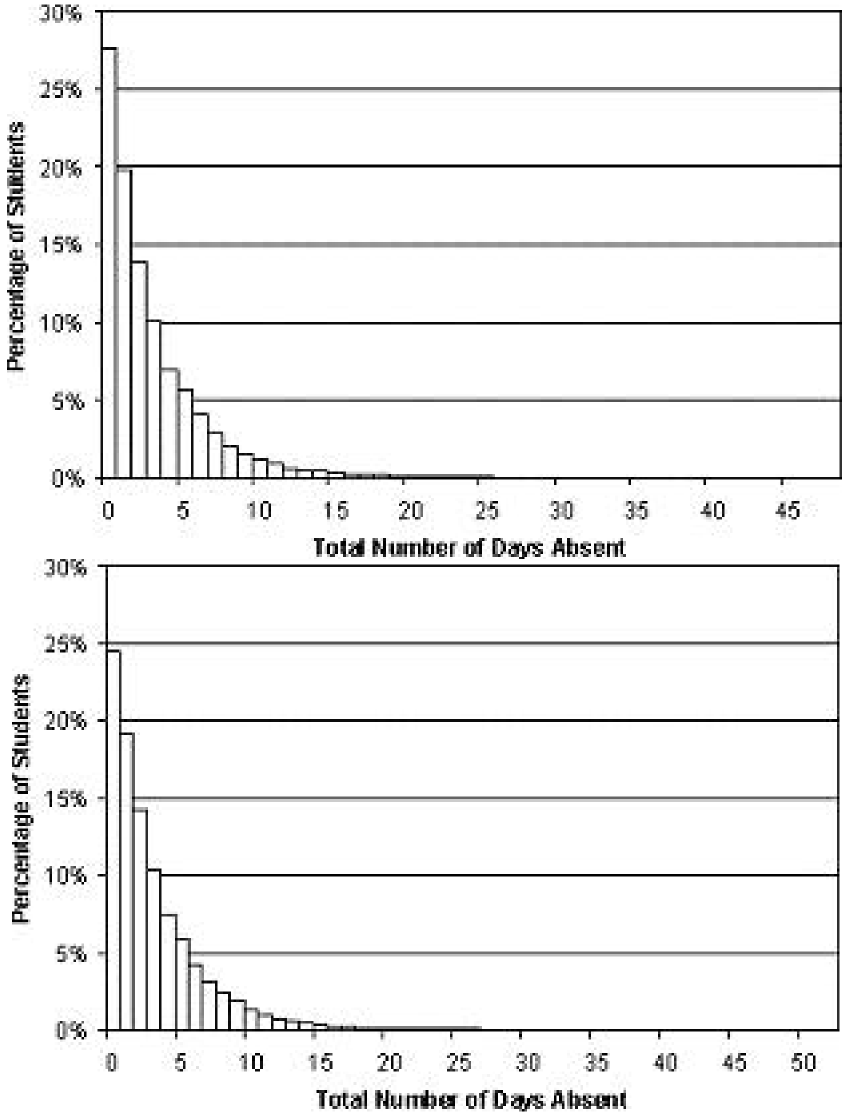

Figure 1. Distribution of number of days absent per student, 1997 (top) and 1999 (bottom)

Analysis of absences across all students in all schools showed that the data did not fit a normal distribution. In 1997, 27.7 per cent of students had no absences all term; in 1999, 24.4 per cent had no absences. The high number of zeros in the data (see Figure 1) meant that a standard transformation could not be used to approximate a normal distribution. HLM offers computational options for a dependent variable that represents counts. The Poisson option in HLM results in a nonlinear analysis using a hierarchical generalised linear model (Bryk, Raudenbush and Congdon 1996, ch. 5). The analysis proceeds adding variables in three stages, with adjustments at each stage to include only significant variables. The final model shows how each of the variables influences a school's absence rate. The analysis was first done using the 1997 data, with 1999 used as a replication. The following discussion considers the 1997 analysis; results for 1999 are contained in the tables. The steps in the analysis are similar to those followed by Rumberger (1995) in his analysis of middle-school dropouts.

Models

The first step in a multilevel analysis is to estimate a model with no student- or school-level variables, estimating the variances in the dependent variable at the student and school levels and testing whether there are significant differences between schools. Transforming the estimate of the intercept in the HLM analysis, the predicted mean of the number of absences per school is 3.02 for 1997. The estimated variance of the intercept term is small (0.091) but significant. The estimated number of absences per student would most likely fall between 2.23 and 4.08 [$\exp(1.104 \pm (.091)^{1/2})$], one standard deviation below and one standard deviation above the mean.

The next step in the analysis was the addition of student-level predictors to the model. As noted above, student characteristics were contained in three dummy variables: SEX, ABOR and CARD, representing gender, indigenous background and socioeconomic status, respectively. Sex was not significant and was removed from the model and all subsequent analyses. Indigenous background and socioeconomic status were both significant, but the estimated variance for SES was small, so that its effect on the overall model was minimal. For subsequent analyses, the slope for SES was fixed; the slope for indigenous background remained random. Once the model was re-evaluated with only significant student-level variables and the SES slope fixed, the estimated parameter variance for the intercept was smaller than the estimated parameter variance in the ANOVA with random effects (0.065 versus 0.091). Controlling for differences in the background characteristics of students accounted for 31 per cent of the variance in average number of absences per school (Table 2, Column 2).

Table 2. Summary of results for variance explained by HLM models, 1997 and 1999

	1997		*1999*	
	Parameter (1)	*Variance Explained (2)*	*Parameter (3)*	*Variance Explained (4)*
Mean number of absences per student (β_0)				
Intercept (δ_{00})	1.104		1.195	
Variance (τ_{00})	0.091**		0.090**	
Student-level variables only	0.063**	30.9%	0.062**	31.2%
+ Location	0.061**	32.7%		
+ Student composition	0.056**	37.8%	0.055**	38.9%
Indigenous background (β_2)				
Intercept (δ_{20})	0.471			
Variance (τ_{22})	0.381**			
+ Student composition	0.349**	8.5%		

** Significant at .01 level.

School-level variables were then added, first by including school location (metropolitan or country). Although children live in country or metropolitan areas, it was decided that location better described the school rather than its students. There are many examples of country students enrolled in metropolitan schools, especially because of the classification of schools. The Australian Bureau of Statistics (ABS) classifies statistical local areas (SLAs) as metropolitan if they fall within a boundary marked by Gulf St Vincent in the west, Gawler in the north, the Adelaide Hills to the east, and Aldinga to the south; Stirling, in the southeastern hills, is also considered metropolitan. All other parts of the state are considered non-metropolitan (country), including communities well within commuting distance to Adelaide. The addition of location was significant, although it reduced the estimated parameter variance of the intercept by only an additional 1.8 per cent. It had no effect on the estimated parameter variance of the student-level variable ABOR.

Student composition of the school was the next set of variables to be added to the model. This set comprised three variables, PCTABOR, PCTCARD and SIZE, which were added simultaneously to the intercept and the student-level variable ABOR (CARD was fixed). Only PCTABOR and PCTCARD had significant effects on the intercept, but only PCTABOR was significant on the ABOR slope. The school size variable, SIZE, was not significant. The addition of these variables reduced the parameter variance of the intercept an additional 5.1 per cent, for a total of 37.8 per cent. When PCTABOR was added to the slope of ABOR, that parameter's variance was reduced by 8.5 per cent.

The final model for the 1997 data was

$$\text{TOTABS} = \beta_{0j} + \beta_{2j}(\text{ABOR}) + \beta_{3j}(\text{CARD}) + \varepsilon_{ij}$$

at the student level, and

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{LOCATION}) + \gamma_{02}(\text{PCTABOR}) + \gamma_{03}(\text{PCTCARD}) + u_{0j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{22}(\text{PCTABOR}) + u_{2j}$$

$$\beta_{3j} = \gamma_{30}$$

at the school level.

The procedures were replicated for the 1999 data, with both ABOR and CARD remaining as the only significant student-level variables with the CARD slope fixed. The first school-level variable, LOCATION, was not significant for 1999, and was rejected. Among the school composition variables, only PCTCARD had a significant effect on the intercept; no variables affected the ABOR slope. The addition of student-level and school-level variables to the 1999 model reduced the parameter variance of the intercept by 38.9 per cent, more than the 1997 model and with fewer explanatory variables. The final model for 1999 was

$$\text{TOTABS} = \beta_{0j} + \beta_{2j}(\text{ABOR}) + \beta_{3j}(\text{CARD}) + \varepsilon_{ij}$$

at the student level, and

$$\beta_{0j} = \gamma_{00} + \gamma_{03}(\text{PCTCARD}) + u_{0j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$\beta_{3j} = \gamma_{30}$$

at the school level.

Results

The models presented above have shown that differences in school absence rates—as represented by the mean number of days absent per student—are affected by school location (in 1997) and student composition. Within schools, an individual student's absence rate is affected by indigenous background and low SES background. The final estimates of the effect of each of the student- and school-level variables are contained in Table 3. The estimates in the table represent the multiplier effects at student and school levels.

In the final 1997 model (Table 3, Column 2), school location made a difference, with metropolitan schools having a mean absence rate about 8.2 per cent lower than country schools, controlling for the per cent of students in the school who are indigenous or from low SES families. That is, for a school in Adelaide, the school absence rate would be 91.8 per cent of the rate of a country school with the same student composition. While this is statistically significant, across an entire school year it is equivalent to approximately one full school day. Considering that the data were collected in Term 2, which is the only term with Monday holidays, this difference may be explained by absences for "family/social" reasons—most likely for travelling over a long weekend In 1999, there was no difference by location (Table 3, Columns 3-4), when there were three long weekends.

The school mean absence rate was also influenced by the percentage of students of indigenous background and from low SES families. For each per cent of the total school population who were from indigenous background, the mean school absence rate increased by 0.5 per cent. For example, a school's absence rate increased by 10 per cent if it had an indigenous percentage of students 20 per cent higher than another school. A school with 50 per cent indigenous students would have an absence rate 10 per cent higher than a school with 30 per cent indigenous students.

Table 3. Estimates of adjusted school mean absence rates, indigenous background and socioeconomic status, primary students

	1997		1999	
School-level variables	*(1)*	*(2)*	*(3)*	*(4)*
Mean absence rates [a]				
Location (Metropolitan schools)	0.916**	0.918**		
Per cent indigenous		1.005*		1.004**
Per cent low SES		1.002*		
Indigenous background [b]				
Intercept		1.602**		1.622**
Per cent indigenous		1.019**		
Low socioeconomic status [c]				
Intercept		1.193**		1.246**

* Significant at .05 level.
** Significant at .01 level.
[a] Coefficients represent the estimated effects on the mean absence rate due to a one-unit change in the listed variable. Estimates control led for student-level variables of indigenous background and low socioeconomic status.
[b] The intercept term represents the estimated effect for indigenous background, while the other coefficient represents the change to the intercept for a one-unit change in the listed variable.
[c] The intercept term represents the estimated effect for low socioeconomic status, which was fixed across all schools.

Indigenous students, on average, had higher absence rates than non-indigenous students. The intercept for indigenous students was allowed to vary according to the percentage of indigenous students in the school, and this effect was significant. In a school with the mean percentage of indigenous students (3.2%), an indigenous student's absence rate was 60 per cent higher than the absence rate of a non-indigenous student. For a school with an above-average percentage of indigenous students, an indigenous student's rate increased by 1.9 per cent for each per cent above the mean indigenous population; for a school with a lower-than-average percentage of indigenous students, an indigenous student's absence rate was less than 60 per cent higher than a non-indigenous student's rate. In 1999, no school-level factors affected the student-level factor of indigenous background.

While this shows that indigenous students, on average, were absent more frequently than non-indigenous students, it also shows that a student's indigenous background has less impact on non-attendance than first thought. Summary data in Table 1 showed that the absence rate for indigenous students was 2.6 times the rate for non-indigenous students in 1997, and 2.4 times in

1999. The estimates from the HLM analysis show that when other factors are controlled—such as school location and the percentage of indigenous and low-SES students in the school—the effect on the absence rate decreases to 1.6 in both years. Indigenous students still miss on average more than 60 per cent more school than non-indigenous students—equivalent to about 7 school days across an entire year—but because this figure is much smaller than originally calculated using simple means, more reasonable targets can be set for absence reduction programs for indigenous students.

In both 1997 and 1999, a lower SES student had a 20 per cent higher absence rate than a middle/upper SES student. Thus, if a middle/upper SES student had missed 5 days during the term, a lower SES student in the same school would have missed 6. The simple means showed differences of about 33 per cent; but controlling for student- and school-level factors, this difference is reduced. Although there were indications of interaction between indigenous background and low socioeconomic status, no significant effect was found to be significant.

Discussion

Absenteeism is believed to have a major impact on student learning, but just how absenteeism affects academic achievement has not yet been explained. The simple examination of indigenous students' absence rates presents a serious challenge to educators in Australia, even if there are other factors that influence what appears to be higher rates. While indigenous students' absence rates are not as high as first thought, they are still higher, on average, holding other factors constant, than non-indigenous students' absence rates by about 60 per cent. Similar findings exist for students from lower SES backgrounds: lower SES students' absence rates are higher than middle/upper SES students', but the difference is not as great after controlling for school-level factors.

The finding reported above about the percentage of indigenous students in a school and its effect on an indigenous student's absence rate gives credence to theories stating that educational disadvantage is exacerbated by concentrations of similarly disadvantaged students, although this applies in this analysis to indigenous students only. Is it an issue of relevance for indigenous students? Schools that enrol higher proportions of indigenous students are located in more remote areas of South Australia, and these schools do have higher absence rates than other schools.

While differences in absence rates vary according to student background, this explains only 38 to 39 per cent of the variance; much of the variance in school absence rates remains to be explained. Some of this variance may be explained by student factors not used in the present study, such as those that examine attitudes toward school, parents' education levels and previous achievement, among others. School factors that were not included, such as school organisation, leadership and age of the teaching staff, may also help to explain some of the variance (see Bryk and Thum 1989; Bos, Ruitjers and Visscher 1992; Corville-Smith, Ryan, Adams and Dalicandro 1998).

This study has highlighted the importance of choosing an appropriate design for analysis of school data, especially when the data are gathered as part of an educational system's administrative collection. Such data are often used to establish simplistic benchmarks for the system, and for each individual site within the system, as part of an accountability program. While such an approach may be the ideal because we believe that student background should not have a negative influence on student achievement, the reality is that there are still achievement differences associated with background characteristics. If benchmarks for attendance are to be set, they must account for some of the differences between student composition of the school, otherwise schools may be undeservedly penalised.

References

Altenbaugh, R. J., *et al*. (1995) *Caring for kids: a critical study of urban school leavers*. Falmer Press, Bristol, PA.

Australian Bureau of Statistics. *Schools, Australia* (Catalogue No. 4221.0).

Tj. Bos, K., Ruitjers, A. M. and Visscher, A. J. (1992) Absenteeism in secondary education. *British Educational Research Journal*, 18, 381-395.

Bryk, A., Raudenbush, S. and Congdon, R. (1996) *HLM: hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs*. Scientific Software International, Chicago.

Bryk, A. S. and Thum, Y. M. (1989) The effects of high school organization on dropping out: an exploratory investigation. *American Education Research Journal*, 26, 353-386.

Corville-Smith, J., Ryan, B., Adams G. and Dalicandro, T. (1998) Distinguishing absentee students from regular attenders: the combined influence of personal, family, and school factors. *Journal of Youth and Adolescence*, 27, 629-640.

deJung, J. K. and Duckworth, K. (1986) Measuring student absences in the high schools. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April [ED 271 889]

Fernandez, R. R. and Velez, W. (1989) *Who stays? Who leaves? Findings from the ASPIRA five cities high school dropout study*. ASPIRA Association, Washington, DC.

Finch, F. H. and Nemzek, C.L. (1935) Attendance and achievement in high school. *School and Society*, 41, February 9, 207-208.

House of Representatives Standing Committee on Employment, Education and Training (1996) *Truancy and exclusion from school*. AGPS, Canberra, p. 3.

Kersting, J. (1967) Absences and averages. *School and Community*, 53, February, 17.

Odell, C. W. (1923) The effect of attendance upon school achievement. *Journal of Educational Research*, 8, December, 422-432.

Reid, K. (1982) The self-concept and persistent school absenteeism. *British Journal of Educational Psychology*, 52, 179-187.

Rothman, S. (1999) Non-attendance and student background factors. Paper presented at the joint annual conference of the Australian Association for Research in Education and the New Zealand Association for Research in Education, Melbourne, November-December.

Rumberger, R. (1995) Dropping out of middle school: a multilevel analysis of students and schools. *American Educational Research Journal*, 32(3), 583-625.

Wright, J. S. (1978) Student attendance: what relates where? *NASSP Bulletin*, 62, February, 115-117.

Acknowledgment

The author gratefully acknowledges assistance from Professor John Keeves of The Flinders University of South Australia, and Margaret Ford, Shirley Dally, Dan Turner and Gary O'Neill of the Department of Education, Training and Employment. All content, however, is the responsibility of the author.

IEJ

[1] A *non-attendance* occurs when a student is not present at the school. For some non-attendances, such as school sport, camps and excursions, and work experience, students are considered present, although at a different site. A student is considered *absent* for the following reasons: illness, family/social activities, exemption, suspension and exclusion without an alternative program. An unexplained non-attendance is also considered an absence.

[2] Socioeconomic status is measured according to whether the student is recipient of a "school card," based on family income. A school card entitles a student to subsidies for school fees and other school-related expenses. This variable is dichotomous: a student either receives or does not receive a school card. Students with special needs enter into a negotiated curriculum plan; a student either has a plan or does not have a plan.