DEVELOPMENT OF AN ONLINE PSYCHOMETRIC TEST OF SPATIAL ABILITY

Kenneth John Sutton

Master of Educational Studies
Bachelor of Educational Studies
Graduate Diploma in Educational Studies
Diploma in Teaching (Industrial Arts)
Engineering Trade (Patternmaking)

Submitted in the fulfilment of the requirements for the degree of

Doctor of Philosophy (Psychology)

School of Psychology
Faculty of Science and Information Technology
The University of Newcastle, NSW Australia

February 2011

ABSTRACT

This thesis reports on the development of an online psychometric test of spatial ability for designers that measures choice accuracy and reaction times. The research identified five spatial skills that collectively contribute to a construct called spatial ability. The test consists of 20 test items divided equally into five subtests where each measures a separate spatial skill. The five spatial skills were appropriately named and descriptions of each are provided in the final chapter. Evidence from nine sequential studies based on detailed statistical investigation including item analysis and exploratory factor analysis was used to establish the test. The important psychometric properties of reliability, validity, correlation and effect sizes were constantly assessed throughout the studies, and both parametric and nonparametric procedures were used where appropriate. Participants who took part in the studies were mostly students undertaking courses at university level and were recruited from both design and nondesign disciplines. Sample sizes for the different studies varied, but reached 650 in a final study consisting of male and female participants spread across 15 design disciplines. Two versions of the test were developed and both provide instructions and feedback to the test taker, and participation is possible without the assistance of a supervisor. One version is meant to be used by novice designers or instructors for diagnostic purposes, while the second collects data and demographic information and is intended for research. This thesis established the importance of spatial ability in a design environment and the need for a specific test for designers. Other outcomes include gender differences, the opinion of subject matter experts, comparison between design and nondesign groups and the impact of practice effect on the assessment of any real learning that may occur in a classroom setting. Two methods of item analysis were applied to appropriate datasets, and the relationship between spatial ability and general academic ability was investigated. Choice accuracy and reaction time data were analysed, and the studies mostly report quantitative research, though some qualitative research is also reported. This research examined a large number of subtests and test items that were reduced to the final configuration after strict compliance to psychometric test development standards. Both laboratory and online studies were conducted to help achieve the final outcome.

STATEMENT OF ORIGINALITY

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

Signed: ________________________________

DEDICATION

The completion of this thesis is dedicated first and foremost to my loving wife Yvonne who has been my best friend, confidant and companion since I was 17 years of age. Yvonne is truly the most important human being in my life whose unquestionable support and devotion made impossible tasks achievable throughout our life together. Yvonne has been an essential partner throughout my journey of mature-age education from those early school days through to the present time. Always doing things to help, always understanding and always sharing the determination and sacrifice required. I owe Yvonne the greatest thank you of all time for her love, encouragement, confidence in my capacity to succeed, and for overlooking the many things I neglected along the way. Without Yvonne's love, this PhD endeavour would have no meaning beyond a mere academic achievement.

A second dedication is to my loving parents Allan and Isabel. Both are now in their 90s who never had the opportunity for higher education because of the economic and domestic difficulties they experienced during their adolescent years. Despite their denied opportunities, they nevertheless saw the value of education and did everything possible to encourage their four children to seek the security and satisfaction often associated with higher education. To them, I owe a deep sense of gratitude for being the parents they were, for providing the benefits of a loving, safe and secure home, and for giving me a start that they themselves never had.

ACKNOWLEDGEMENTS

My first acknowledgement is to my loving wife Yvonne for always being there when I needed her most, for making the many sacrifices that were imposed upon her, and for taking care of the many things I conveniently put aside. Without question, the completion of this PhD would not have been possible without Yvonne's understanding, cooperation and ongoing support.

My second acknowledgement is to my primary supervisor Professor Andrew Heathcote for inviting my PhD candidature. I appreciated being guided by Andrew's work ethic, his passion for research, the academic rigour he insisted upon and the learning I received from our discussions and scientific writing he encouraged. My appreciation extends to Andrew's lasting confidence in me as an academic within the School, and also as a PhD student. His continual encouragement and reference to a potentially successful academic career regardless of a late start had a significant influence on my motivation. I thank Andrew for his ageless attitude to postgraduate students, for his grounding in experimental psychology, and for the respect and enthusiasm I developed for the pursuit of new knowledge.

My third acknowledgement is to my second supervisor Dr Miles Bore for respecting me as a colleague, valuing my opinion, his patience in listening to my point of view, his compliments towards many aspects of the draft thesis, and his identification of my developing expertise. Miles was especially helpful because of his effective communication style, his suggestions towards the draft thesis, the priority he gave to me at different times, and his demonstrated professionalism on every occasion. I am also most grateful to Miles for providing lecturing and teaching opportunities that helped me develop a specialisation within psychology. Miles is truly dedicated to the discipline of psychology, and I thank him for the manner in which he approached our discussions, the enrichment he gave to my understanding of psychometrics, and most importantly, the confidence he expressed in my ability to achieve.

My fourth acknowledgement is to Kim Colyvas, the senior statistical consultant commissioned by the Faculty of Science and Information Technology to assist staff and students in their research activities. Kim stands out because of his ability to quickly identify what is concerning the student, his software expertise, his attention to detail and his plain English approach to explaining statistics. His approach and patience are highly praiseworthy, and his viewpoint that there are no rules, expectations or milestones when embarking on a new learning experience was reassuring. I thank Kim for the excitement I have developed for statistics, for embracing my project like his own, and for making it possible to be a student of statistics over a lengthy period of time.

My fifth acknowledgement is to Associate Professor Anthony Williams who appeared to recognise something in me that I did not recognise myself. I will always appreciate Tony's ongoing collaboration, his willingness to co-author many publications, his assistance with recruiting suitable research participants, and his contribution to a successful Australian Learning and Teaching Council (ALTC) grant application. I thank Tony for his long friendship, his continual support, and the career aspirations he always maintained for me.

My sixth acknowledgement is to Associate Professor Don Munro for his willingness to be a reader of my draft thesis. Don achieved this task in a very short period at a very inconvenient time and provided invaluable advice that I adopted to strengthen the standing of the thesis. I thank Don for his thorough reading, his attention to detail, his sound contribution and his readiness to elaborate on any issue I wanted to raise.

I also thank the executive of the School of Psychology for allowing me as a new staff member after a workplace restructure to take on this candidature. I appreciated their confidence in my ability to overcome any disadvantages I might encounter. Thanks also to a succession of Heads of School for their genuine encouragement, and for the meaningful discussions that occurred from time to time. Also thanks to every member of staff who accepted me without qualification despite not having a traditional background in psychology. This gave me every incentive to succeed. Finally, my thanks extend to the Pro Vice-Chancellor of the Faculty of Science and Information Technology for his initiative in setting up a statistical support service to assist staff and students in their research endeavours.

PUBLICATIONS AND GRANTS

Publications

Year 2010

A Rationale for Developing Spatial Skills in a Design Environment
Sutton, K., & Williams, A.
Connected 2010 – 2nd International Conference on Design Education 28 June – 01 July 2010, University of New South Wales, Sydney, Australia

A Multidisciplinary Approach to Measuring Spatial Performance in Novice Designers
Sutton, K., & Williams, A.
International Congress of Applied Psychology 2010 (ICAP 2010), 11-16 July 2010, held at the Melbourne Convention and Exhibition Centre in Melbourne, Australia

Implications of Spatial Abilities on Design Thinking
Sutton, K., & Williams, A.
Design Research Society (DRS) international conference Design & Complexity held July 7-9, 2010 in Montreal (Quebec), Canada, at the School of Industrial Design, Université de Montréal

Year 2009

Improving Spatial Abilities with 3D Mini-Maps
Nesbitt, K., Sutton, K., Wilson, J., & Hookham, G.
The 6th Australasian Conference on Interactive Entertainment , Sydney, Australia 14-16 Dec 2009

Spatial Ability Performance of Female Engineering Students
Sutton, K., Williams, A., & McBride, W.
The Australasian Association for Engineering Education (AAEE) 20th Annual Conference Adelaide, Australia 6-9 December 2009

Exploring Spatial Ability and Mapping the Performance of Engineering Students
Sutton, K., Williams, A., & McBride, W.
The Australasian Association for Engineering Education (AAEE) 20th Annual Conference Adelaide, Australia 6-9 December 2009

Spatial Abilities and its Implication for Novice Designers
Williams, A., & Sutton, K.
The 5th Annual DesignEd Asia Conference, Wan Chai, Hong Kong 1 & 2 Dec 2009

Year 2008

Developing a discipline-based measure of visualisation
Ken Sutton & Anthony Williams
2008 National UniServe Conference 1st October - 3rd October 2008 The University of Sydney

Profiling Spatial Ability Performance of Novice Designers
Anthony Williams, Ken Sutton & Rebecca Allen
The 5th Biennial International Conference on Technology Education Research 27 - 29 November 2008 Gold Coast, Queensland, Australia

Spatial Ability: Issues Associated with Engineering and Gender
Anthony Williams, Ken Sutton & Rebecca Allen

The Australasian Association for Engineering Education (AAEE) 19th Annual Conference Yeppoon, Australia 7-10 December 2008

Year 2007

Measuring 3D understanding on the web and in the laboratory
Sutton, K., Heathcote, A. & Bore, M. (2007).
Behavior Research Methods, 39 (4), 926-939.

Research outcomes supporting learning is spatial ability
Sutton, K., & Williams, A. (2007).
Proceedings of the 2007 AaeE Conference, Melbourne, 9-12 December 2007

Spatial cognition and its implications for design.
Sutton, K., & Williams, A. (2007).
Proceedings of the International Association of Societies of Design Research, The Hong Kong Polytechnic University, 12-15 November 2007

Developing interactive computer-based learning experiences to improve spatial ability.
Williams, A., & Sutton, K. (2007).
Reflection paper, IADIS International Conference on Cognition and Exploratory Learning in Digital Age 2007, Algarve, Portugal, 7-9 December 2007 [E1]

Year 2006

Impact of Spatial Ability on Students Doing Graphics Based Courses
Sutton, K., & Williams, A.
17th Annual Conference of the Australasian Association for Engineering Education Auckland, New Zealand 10-13 December 2006

Performance Differences on Spatial Ability Tasks
Sutton, K.
Australian Conference on Personality and Individual Differences, Newcastle, 1-2 Dec, 2006

Grants

Australian Learning and Teaching Council (ALTC) Grant. Assessing and Improving Spatial Ability for Design-based Disciplines Utilising Online Systems. Agreement dated 19 October 2007 ($196,331).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

TERMS AND DEFINITIONS

3D Ability Test. The online psychometric test of spatial ability which is the central focus of this research is called the 3D Ability Test. The test is also known by its acronym (3DAT).

3D Understanding. Regarded as an alternative name for spatial ability but emphasizes the importance of 3D in the interpretation of spatial ability. 3D understanding is also described as visualizing in three dimensions.

Algorithmic Design. A position at one end of the design spectrum that could be described as structured and restricted. Design that is driven by procedures, formulae and precedents.

Analytic Solution Strategies. This approach to solving problems depends on a systematic step by step strategy. Analytic procedures handle spatial information by breaking it down into nonspatial list-like elements and essentially treat one element at a time. Fewer errors may occur this way, but responses may be slower than alternative strategies.

Classical Test Theory. The traditional and well established approach to item analysis is known as the *classical test theory* model and combines measures of item difficulty, discrimination and reliability to determine the acceptability of test items.

Course. An academic subject that is part of a tertiary education program such as a degree or diploma. Also applies to primary and secondary education as part of an award upon graduation.

Design. A general term that covers disciplines engaged in design processes at any level in a range from algorithmic design to heuristic design. Disciplines include engineering, architecture, mechatronics, construction management, industrial design, surveying and graphic design.

Effect Size. Effect size is a calculation that provides a standardized measure of a difference within a group or a difference between two groups. Effect size could also be described as practical significance and it allows easy comparisons to be made. Standard benchmarks have emerged and are expressed in terms of indexes representing small, medium and large effects. Where standard deviations (SD) are similar, effect size can be calculated using the formula: $(MEAN_1 - MEAN_2)/SQRT((((N_1\text{-}1)*SD_1^2)+((N_2\text{-}1)*SD_2^2))/(N_1+N_2\text{-}2))$. Where SDs differ greatly, the formula used is: $(MEAN_1 - MEAN_2)/SQRT((SD_1^2+SD_2^2)/2)$. The concept of effect size can also be applied to relationships between variables (correlations). Standard benchmark indexes distinguish between small, medium and large effects.

Graphical Communication. Graphical communication generally describes an academic course dedicated to developing communication skills linked to graphics. Includes notational systems, sketching, documentation, computer assisted design (CAD) and the production of technical drawings. Also generally includes some activities requiring spatial ability.

Heuristic Design. Positioned at the opposite end of the design spectrum to algorithmic design and is considered to be unstructured and creative. Design that is not generally driven by set procedures and constraints, but by general principles and objectives.

Holistic Solution Strategies. A holistic approach to solving problems is essentially seen as visualizing and reacting to a task as a whole. This approach is considered more efficient because responses to stimuli generally involve less time.

Item Difficulty. This is a measure of the degree of difficulty of a test item indicated by an index number calculated according to the proportion of test takers getting the item correct. The higher the index number, the easier the test item is.

Item Discrimination. This is a measure of how well an item in a test separates high scorers and low scorers on the total test as a whole. A test item is not performing adequately if high scorers on the test as a whole score poorly on the test item.

Item Reliability. Item reliability is a measure of the internal consistency of test items within a test, or the internal consistency of test items within a subtest. A high index number indicates high internal consistency. Cronbach's alpha is one measure of item reliability.

Item Response Theory. A method of item analysis gaining acceptance is known as the *item response theory* model or sometimes referred to as the *latent trait theory* model. In simple terms, it represents each test item under consideration as a plot on a graph where one axis represents ability while the other represents the probability of a correct answer. The shape of the plot provides a quick assessment of the standard of the test item. There are one, two and three parameter versions of the item response theory model.

Latent Trait. A present but not active or able to be seen distinguishing characteristic or quality of a person.

Program. A program is also known as an undergraduate degree or equivalent offered by a university. Programs consist of courses that typically lead to the award of a Bachelors degree or equivalent after successful completion of the courses.

Reaction Time (RT). A measure taken that accurately shows the time taken for a test taker to respond to a test item or stimulus. It is generally reported in milliseconds. Other terms that may be used are *response time* and *time taken*. Note that RT on occasions in this thesis can also refer to a subtest called *Transformation* and its abbreviation is also RT. The context in which RT is used will make it obvious which meaning is applicable.

Reliability. A measure of reliability which is generally in the form of an index number reports the level or repeatability of results for a test or test item. There are a number of forms of reliability such as internal consistency and test retest reliability. A good test should show high repeatability to be a meaningful instrument. A test can be reliable independent of its validity.

Sociological Experiences. This refers to life experiences especially during developmental years that shape human attitudes, behaviour and personal skills. Many possible life experiences and activities engaged in during formative years are thought to have an impact on the development of spatial skills and understanding.

Spatial Cognition. Regarded as higher level spatial thinking that involves all aspects of related knowledge including: perception, thinking, imagining, reasoning, judging, remembering and communicating.

Spatial Ability. Defined as the performance on tasks that require the mental rotation of objects, the ability to understand how objects appear in different positions, and the skill to conceptualize how objects relate to each other in space. Of note is that one disadvantage of using the word *ability* in a descriptor is that it can have a connotation of something that is less predisposed to change and seen instead to be more innate. For this reason, there would be a preference among many in the discipline to using *spatial performance* rather than *spatial ability*.

Spatial Factors. Spatial factors are known as elements, classifications or components of spatial ability. A spatial factor can also be described as a spatial skill. Spatial ability consists of a number of spatial skills, and collectively they provide an assessment of spatial ability. A different test or a set of tests are required to measure separate spatial factors.

Spatial Tests. Instruments designed to measure spatial ability. For the most part, they consist of test items of single design that are used for a wide range of purposes across a variety of disciplines.

Subject Matter Experts. Professionals from industry employment or academia where spatial ability is a strong requirement for success in that profession. The term is also abbreviated and used throughout as SME.

Subtest. Essentially a spatial test that is one subset of a larger test of spatial ability. A subtest is intended to measure a single specific spatial skill or help identify particular learning difficulties.

Technical Drawing. A technical drawing is a collection of related views of an object (e.g., buildings, machine parts) and a notational system that combine to graphically represent and convey technical and structural information about that object. It is an essential skill for industrial professions and tradespeople. An alternative name is engineering graphics.

Test Item. Each test subtest of any ability is made up of stimuli designed to capture the performance of a test taker. *Test item* is an alternative name for stimulus or question.

Test of Single Design. Test items that make up the stimuli in a subtest are considered to be of a single design. This means that the test items or stimuli are similar in style and shape and they are simply variations of the same design and purpose. In contrast, a test of spatial ability in the context of this research consists of a range of subtests and therefore a mixture of test items.

Test Type. A term used to describe a type of test or a subtest.

University Admissions Index (UAI). A person wishing to enrol in a university degree is often given a score based on their prior academic achievements which is used or partly used to decide if a place will be allocated, especially where places are competitive. In some states in Australia, this score is called a University Admissions Index (UAI). There are generally equivalents to this in other states or in other countries. In many respects, UAI is a good measure of general academic ability.

Validity. The validity of a test is the degree to which it measures what it claims to measure. There are a number of validities and those relevant to this research are defined in the thesis as they appear. These validities are named: content validity, construct validity, convergent validity, discriminant validity, face validity and validity with known groups.

Visualization. The ability to call up or form mental images or pictures or to make perceptible to the mind or imagination. The process of forming mentally visual images of objects not present to the eye.

Visual Perception. The process of selecting, transforming, organizing and interpreting graphical information received through our visual sensory receptors. The ability to interpret what is seen.

TECHNICAL DRAWING CONCEPTS

A technical drawing is defined as a collection of views of an object and a notational system that combine to communicate technical information about that object. Technical drawing is usually based on a standard set of three axes (X, Y and Z) meeting at right angles at a point called the origin. The axes provide a cartesian coordinate system for locating points, lines and planes. Reference planes (also called viewing planes) are usually defined, one parallel to the XY axes (XY plane), another to the XZ axes (XZ plane) and the third to the YZ axes (YZ plane). The details of an object are normally represented by a set of 2D drawings where each represents a different view of the object. 2D views are referred to as orthographic views. To produce a set of orthographic views, an object is theoretically positioned with respect to the three axes and individual views are projected perpendicularly on to each of the reference planes (see Appendix A, Figure A01). This method of projection is termed orthographic projection. The view seen through the XY plane is called the Top View (TV), the view seen through the XZ plane is called the Front View (FV) and the view seen through the YZ plane is called the End View (EV). 2D views produced by orthographic projection are termed degenerate views because one axis is excluded. The distance of the object from the reference planes is not critical to the shape of the projected views as the shape is the same regardless of the distance in an orthographic projection (Sutton, Heathcote, & Bore, 2007). This is a convention that contrasts with perspective drawings where the distance behind the viewing plane is critical to what is seen.

There are three important concepts that are fundamental to orthographic projection. These are termed *true shape*, *true length* and *true angle* and all three are related. Essentially they imply that the true shape of a surface, the true length of an edge and the true angle of inclination of an edge or surface are not always seen in a 2D view. Instead, what is often seen are apparent shapes, apparent lengths and apparent angles of inclination. A true angle for example, is seen when a projection of a line or an edge is parallel to a viewing plane. Knowing the difference and the conditions under which true measures are seen is important to the understanding of a technical drawing.

Isometric drawings, which provide more obvious information about the 3D properties of an object than orthographic projections, are produced by a view that is not parallel to any axis, with the view being projected onto a plane perpendicular to the viewing direction. For an isometric drawing, the top view of a line representing the viewing direction is typically at 45°, 135°, 225° or 315° to the X axis. In the standard setting of most Computer-Assisted Design software the viewing direction will be a true angle of 35.3° to the XY plane. In isometric drawings, parallel edges running away from the viewing plane are always drawn as parallel lines. This contrasts with perspective drawings where parallel edges that run from the viewing plane are drawn as converging lines. However, for relatively small objects (as opposed to, say, landscapes) the difference between isometric and perspective drawings is negligible (Sutton et al., 2007). Given this, and the dominant use of isometric rather than perspective as 3D representations in technical drawing, the focus in this thesis is on isometric.

A common approach to developing technical drawing competency is to experience both 2D and 3D representations concurrently in order to develop an understanding of the relationship between the two. It is not desirable to develop 2D or 3D skills in isolation and in many respects, working in 2D may be more important than working in 3D. James et al. (2001) report that participants in their experiments spent more time looking at the end and front views of objects rather than three-quarter or intermediate views. They suggest that these are the views where there is the greatest amount of difference in the visibility of object features. In contrast, the three-quarter views are perceptually similar. The process of working from 2D to 3D drawings, and working from 3D to 2D drawings, is the common way students build up their understanding of concepts. The ability to interpret a multi-view drawing is learnt by forming mental images from the 2D views and visualising what the object will look like in 3D. As the complexity of objects increase, extra views are generally necessary, including sectional views (planes cutting through objects), exploded views (magnified projections showing individual parts separated), and assembled views (working parts in position) (Sutton et al., 2007). The number and type of views to form a technical drawing will depend on how complex the object is and how much information needs to be communicated. Collectively, the 2D views convey precise details about an object intended for manufacture or machining. Technical drawings generally consist of a set of 2D drawings (views) because complex information is more easily represented this way. As a consequence, an object is less frequently drawn as a 3D representation and the visualisation of the object relies heavily on the ability to interpret a set of 2D views. This skill is regarded as being able to *read* a drawing. As touched on earlier, Salthouse (1991) leaves little doubt about the importance of being able to read a drawing. He states that the ability to understand a technical drawing is vital because technical drawings are necessary in the process of converting from a design concept to a physical structure of some form. He adds that the correct interpretation of a technical drawing is critical because they often serve as legal documents to indicate what will be constructed. Costly delays and litigation are possible when errors occur. While designers may not always be engaged in producing actual drawings, Salthouse considers graphical comprehension will nevertheless be an important factor throughout their careers.

PROLOGUE

The research reported in this thesis is about developing an online psychometric test of spatial ability for designers. The name given to this test was the 3D Ability Test (3DAT) and the final version consists of five subtests with four test items in each. These numbers fluctuated throughout the different studies and at one stage 25 subtests and 119 test items were being considered. Each subtest measures a different spatial factor and collectively they assess a construct called spatial ability. From the outset, the prime objective was to establish a 3D ability test for designers (e.g., architects, engineers) with a longer term objective of developing learning tasks to improve spatial ability for novice designers. This research demonstrated an example of applied psychology, and for many studies, a psychology/ design nexus was crucial to outcomes. In many respects, the strengths of the psychology discipline such as expertise in experimental design, psychometrics and statistics were exported to an external discipline to achieve research objectives beneficial to that discipline. In this case, the discipline was design which encapsulated mostly engineers, architects and construction managers in this research. The decision to develop the 3DAT as an online test was based on the easy access it would provide to institutions wanting to use it as part of their curriculum. There were other advantages such as the ease at which changes could be made, the customisation possible, the simplicity in managing, collecting and collating dependable data, and the potentially large samples that could be captured for educational and research purposes. To achieve the online version, a procedure was put in place that included several studies conducted in a lab environment using local software to evaluate the viability of going online with the 3DAT. It was a progressive undertaking that considered such things as timing issues, server implications, data collection, graphical displays and the testing of many variables. The lab studies provided benchmarks that the online version could be compared with during the different stages of development. The data collected from the different studies were mostly quantitative that measured choice accuracy (correctness of answers) and reaction times (RT). The accuracy data were always utilised, but the RT was not reported for every study. The 3DAT was never promoted as a speeded test since accuracy was always seen to be more important than speed. RT was therefore considered difficult to utilise on every occasion. However, RT data proved to be far more meaningful than first thought and was fully exploited in a major study towards the end. The collection of some qualitative data did take place in two particular studies during the later stages of development and they provided meaningful information from a different perspective on both occasions.

The first chapter establishes the importance of spatial ability and provides reasons why the development of the 3DAT was justified in view of the large number of so called spatial tests that are available. The existence of spatial ability elements which are generally referred to as spatial factors is a disputed issue in this field of research. Considerable effort therefore was spent throughout this research to test for spatial factors. The identity of these and the subtests that would measure them were important issues because they primarily dictated what the profile of the 3DAT should be to adequately measure spatial ability.

Because *technical drawing* is an essential tool for designers for communicating technical and sometimes complex information to others, some time was dedicated earlier to explaining the fundamental concepts and the notational system used in technical drawing. The requirement to read a technical drawing and to understand its complexities goes to the heart of why spatial ability is such a central attribute for designers.

The steps in developing the 3DAT were sequential across nine studies and psychometric properties were investigated in different degrees at various stages. For example, initially in a preliminary context, then later in an expanded capacity, or repeated when sample sizes increased. The later stages made it possible to produce more convincing evidence in support of the 3DAT. Studies varied in design and in some cases they were exploratory, and in other cases they focused on specific things such as item analysis, factor analysis, expert opinion and attempts to identify spatial factors. Many participants in the studies were novice designers who came from a range of design fields of study. It was also necessary to include participants from disciplines other than design because they were needed to help demonstrate particular psychometric properties. Every study involved both male and female participants, but males were always in the majority. This condition alone was a reminder of the shortage of females attracted to the design discipline. A substantial part of the investigations examined the all important concepts of validity, reliability and correlation. The diversity of the studies meant that a range of statistical procedures could be applied, and they collectively provided the evidence that the aims of the research had been achieved.

A significant proportion of this research was dedicated to exploring gender differences. The literature consistently reports a difference in spatial ability and a difference favouring males is generally found. However, some studies refute this position and argue that factors that influence findings include the type of test used to measure performance, gender stereotyping, sociological experiences and the type of training (if any) that has occurred. Because gender is generally a factor in most studies related to spatial ability, and because it is an issue many researchers are interested in, it would have been remiss not to have given it due consideration in this research. Consequently, gender issues received coverage from a number of perspectives. Comparisons were made between a design group and a nondesign group, also within those two groups, between disciplines and across various subtests. Findings were mixed, but generally males out performed females, though not in all cases. Sample sizes for female cohorts were not always ideal, but this was a reflection of the state of the design disciplines in general, except say for architecture where numbers were reasonably balanced. This research supported the position that gender difference is not reducing in terms of test scores, though one small study suggested otherwise. Also that the difference is not always robust, and that the group that test takers belonged to had an impact on that difference.

The principles of test construction were applied to the development of the 3DAT, and although these are described in a number of ways and in different levels of detail in the literature, they essentially come down to several fundamentals. For example, a need for the 3DAT was established, and this included firm ideas about what it should measure and what method of measurement would be used. Test item preparation followed which was a lengthy process because of the uncertainty about the design of test items, their psychometric properties, the number required and item variation. Next the 3DAT was subjected to a series of pilot testing and several features of the experimental design were revised accordingly. Item analysis was somewhat continuous throughout developmental stages and resulted in various degrees of item reworking, item replacement, layout changes and the retesting of psychometric properties. The more indepth side of item analysis where subtests and test items could be permanently discarded occurred in the latter stages of development after large sample sizes had been achieved. Consequently, this achievement allowed full confidence in the analysis undertaken. This research conformed to commonly held views of test construction principles which essentially became a set of guidelines that pervaded every stage of development. The position taken was that accurate tests do not simply happen, but instead, they are the result of a systematic approach to test development procedures established over time.

FORMAT AND STRUCTURE

This thesis consists of a foreword section, five chapters, a reference list and a set of appendices that are referred to in the chapters. The appendices serve a more significant purpose in this thesis than may be usual since they are seen as an alternative to adding large Tables and Figures to the main text. They also provide additional information that complement the key facts reported in the thesis which a reader may find of special note. Ideally, the appendices will prove to be convenient and will be reviewed alongside the main text when referred to. Chapter 1 introduces the spatial ability construct and examines concepts, issues and theories that all underpin spatial ability in some way. A literature review is part of this chapter, but it is not developed as a separate section. Instead, it is integrated with other reporting with the intent of supporting, challenging or clarifying points as they arise. Chapter 1 also provides a rationale for conducting the research and presents hypotheses to clearly state the direction of the research. Also included towards the end of this chapter is a brief overview of the stages devoted to developing the 3D Ability Test (3DAT) since this development was the fundamental objective of this research. Chapter 1 concludes with a brief description of the final 3DAT and a mention of its special features and qualities.

Chapters 2, 3 and 4 are dedicated to specific stages in the development of the 3DAT and they are titled: *initial*, *transitional* and *final* respectively. These chapters are sequential and each covers several studies where each study has a different focus. All three chapters have intentionally similar formats and start with the aim of the study followed by specific objectives, methodologies, analyses undertaken and results. Further, each chapter finishes with a summary of all study outcomes with major findings particularly addressed. Throughout these chapters, psychometric test development principles are paramount and the essential properties of validity, reliability and those revealed from item analysis are foremost in every consideration. As a consequence, a range of statistical procedures are also reported, and coverage is progressive in the sense that issues such as reliability are not examined in every study, but only where required. Sometimes these properties are revisited as part of that progression.

Chapter 5 is concerned with *discussion and conclusions* for the entire research and draws attention to outstanding matters, implications, notable issues and recommendations. In particular, the research hypotheses are evaluated against final outcomes. In simple terms, the evidence from this research is reviewed and particular aspects are reported.

One final point is that gender is a significant topic in the literature and it is a major consideration in this research. In many respects, gender is deserving of a dedicated section, however, the approach taken was to treat this topic contextually where it seemed appropriate to emphasize particular points. This means that gender issues surface in many of the studies, and reporting is sometimes part of the reporting of other findings.