THE UNIVERSITY OF MANITOBA

A Study of Ranking Procedure for Evaluating Off-White Fabrics

by

Chengyu Yang

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfilment of the Requirements for the Degree of
Master of Science

DEPARTMENT OF CLOTHING AND TEXTILES

Winnipeg, Manitoba

September, 1993

ISBN 0-315-86107-X

Name CHENGYU YANG

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

CLOTHING & TEXTILES | 0386 | U·M·I

SUBJECT TERM | SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

A STUDY OF RANKING PROCEDURE FOR EVALUATING OFF-WHITE FABRICS

BY

CHENGYU YANG

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

To: My parents, brothers and sisters

Acknowledgements

Sincere appreciation is expressed to my major advisor, Dr. W. Pelton, who contributed his guidance throughout the study. Special gratitude is extended to Dr. Lena Horne, who contributed her time, ideas and critiques; Dr. K. Mount, who established the statistical model for this study, and strongly supported the study; Dennis Murphy from the Statistical Advisory Service, who brought the threshold test and other statistical ideas to my attention, and contributed time, energy, and valuable assistance.

Specific gratitude is given to Dr. M. King, who provided references and ideas for the theory of colour measurement; Dr. Douglas Broadbent from Sherbrook University, who gave me a starting point to generate fabric specimens with different degrees of whiteness in similar hue; Brigitta Badour and Judy Thoroski, who gave their suggestions, encouragement, and cooperation in the study. Special thanks are extended to the people who participated in and supported the study.

A special thank-you is also given to my family and friends for their concern and encouragement in this undertaking.

Abstract

The researcher investigated the accuracy of a visual ranking method to assess colour differences in off-white fabrics under a series of controlled conditions. Three research questions were proposed relating ranking accuracy to: (a) the magnitude of the colour difference between consecutive specimens, (b) the number of specimens being ranked, and (c) the level of a subject's visual perception.

Twenty-four female students in Clothing and Textiles from the faculty of Human Ecology completed the threshold and ranking tests during the 1993 spring session. The threshold test established the ΔE values for the colour difference factor in the ranking test. Interchanging the specimen positions in the threshold test improved the accuracy of visual perception. Visual ranking results were analyzed using a randomized complete block design under two factors: (a) ΔE between consecutive specimens within a given series ($\Delta E_{min} = 0.2$, $\Delta E_{med} = 0.5$ and $\Delta E_{max} = 0.8$) and (b) number of specimens being ranked at one time (3, 6, 9 and 12).

Kendall's Tau correlation coefficient was calculated to determine the strength of relationship between each subject's ranking order and the expected ranking for each series within both factors. Page's test for ordered alternatives rejected the first two null hypotheses, supporting the alternative hypotheses that ranking accuracy increased as the ΔE between consecutive specimens increased, and decreased as the number of specimens being ranked increased. The overall strength of relationship between visual and instrument rankings illustrated that the median $\hat{\tau}$s ranged from 0.73 to 1.00 when ΔE was equal to or greater than 0.2, and when the number of specimens being ranked varied from 3 to 12. The number of specimens being ranked at one time and the subject's ability to detect minute colour difference influenced ranking accuracy; however, the method of data collection did not affect the median $\hat{\tau}$ values.

The third research question "Does a subject's level of perception affect ranking accuracy?" was not addressed because there were not sufficient observations for statistical analysis; more threshold comparisons per subject were needed.

Table of Contents

List of Tables

Table

List of Figures

Figures

CHAPTER 1

INTRODUCTION

The whitest textile product is judged by consumers as the most desirable. It is considered to be the freshest, cleanest and free from contaminants (Hunter, 1975). The appearance of a product is often an important attribute on which consumers base their purchase decisions (Meilgaard, Civille, & Carr, 1987). Jakobi and Lohr (1987) reported that over the last two decades in America, 40% of the textile purchases were white. According to Hunter (1975), there are about 5000 white shades and 30,000 "ish" white. Studies have illustrated that consumers prefer a bluish-white hue (Hunter, 1961). Consumers of detergents expect a white textile product to retain its original colour (lightness, chroma, and hue values) throughout the product's life whether the product is soiled from normal wear or is stained from unexpected sources. Therefore, detergent manufacturers are interested in assessing colour, colour change and/or colour difference, especially whiteness.

Detergent manufacturers modify formulations and continually monitor consumer acceptance of their product. Assessment of whiteness can be made by objective instrument measurements or by subjective panelist evaluation, which may include paired comparisons or ranking methods. Of the four Canadian detergent manufacturers who were contacted by the researcher in 1991, three reported both types of assessment. All three employed instrument assessment at some stage of product development; two used paired comparison techniques; one also utilized a ranking approach some of the time, but preferred to use paired comparisons. The manufacturer who used only instrument measurement indicated that paired comparison techniques were not appropriate when more than three products were being tested. The four manufacturers contacted did not indicate why the ranking method was not used at the later stages of product development.

The purpose of this study is to investigate the accuracy of a visual ranking method in assessing colour change in off-white fabrics under a series of controlled conditions.

Justification

For proprietary reasons, organizations such as detergent manufacturers are unwilling to share with the public their accumulated research knowledge on sensory and instrument assessment. Sensory evaluation is usually conducted for quality assurance and new product development (Larmond, 1977) and assists management in making business and marketing decisions. Business decisions should be based on a thorough understanding of the physical and chemical factors behind the attributes of interest, as well as on the perception of human beings (Meilgaard et al., 1987). In practice, product development, consumer acceptance and consumer preference are the principal applications of sensory evaluation.

Management is continually seeking cost efficient methods to evaluate products. Ranking, one of the sensory evaluation methods, is less time-consuming and needs less panel training than paired comparisons when visually assessing more than two specimens (Meilgaard et al., 1987). In fact, textile research investigating visual evaluation methodologies such as paired comparisons, triangle tests, scaling, or ranking techniques under a set of controlled experimental conditions has not been located in the literature.

Studies from textiles and home economics sources over the past 30 years illustrate that researchers have used several instrument values to assess colour change or colour difference. Whiteness Index and Delta E have been recommended as standard instrument measurements by both American Association of Textile Chemists and Colourists (AATCC) and American Society for Testing and Materials (ASTM) (AATCC Test Method 110, 1989; AATCC Test Method 153, 1985; ASTM D2244, 1989; ASTM E313, 1987). However, correlations between panelists' ranking and instrument values now used as standards have not been located in textile studies.

In sensory evaluation of food, Buchanan, Givon, and Goldman (1987), and Meilgaard et al. (1987) indicated that ranking is less sensitive than paired comparison because of the memory factor. This limitation may not be valid for visual sensory evaluations of textile products where a panelist is allowed to manipulate the individual specimens within a given ranking series. This ranking approach creates a series of mini paired comparisons in which panelists rank specimen series by assessing adjacent pairs from the whitest to the least white.

In textiles, several researchers reported paired comparison or ranking methods when assessing fabric whiteness by panelists. Some of the researchers compared the results between instrument measurement and visual evaluation (Furry, Bensing, &

Johnson, 1961; Coppock, 1965; Rhode Island Section, 1966; Morris, 1970; Warfield & Hardin, 1981; Galbraith, Swartzlander, & Hardin, 1987).

In textiles, Coppock (1965) pointed out a weakness of the ranking method, when he requested panelists to rank a series of 30 white specimens from whitest to least white, he suggested that the panelists were unable to rank the specimens from whitest to least white because the colour difference between adjacent specimens was too small. He then used a paired comparison method and was able to reproduce a visual ranking similar to that produced by colour measurement instrument. This early attempt at ranking suggested that ranking may be less accurate than a paired comparison method. However, Coppock did not indicate the magnitude of instrument readings which described "small colour differences", and did not test the influences from the changes in the number of specimens being ranked at one time.

Other researchers in textiles supported the use of the ranking method. Furry et al. (1961), Rhode Island Section (1966), and Warfield and Hardin (1981) who conducted ranking procedure on textiles reported "very good" agreement between instrument measurement and visual evaluation. However, they did not use Whiteness Index or Delta E measurements in their studies.

The literature revealed that both instrument measurement and sensory evaluation have advantages and disadvantages. Instrument measurement is far more sensitive than a panelist's visual evaluation of colour change and colour difference; it is quicker and cheaper to use. The trend in assessment of fabric visual attributes appears to favour increased use of instrumentation (Vanderhoeven, 1992). Instruments are designed to simulate human visual perception, but unable to replace human visual responses completely. When instruments detect a colour change, the human eye may not be able to perceive the difference, or the consumer may accept the difference.

Sensory evaluation is required to demonstrate validity of instrument assessment and to standardize objective tests (Larmond, 1977). Visual evaluation can be used to test human responses directly. Product changes can be assessed using visual evaluation when there is no adequate instrumentation; however, visual evaluation is less sensitive and more expensive than instrument measurement.

From the literature review, the researcher observed gaps in the current knowledge and understanding of ranking procedures to assess visual attributes of textile fabrics. These can be summarized as: (a) ranking has been criticized by some researchers as inaccurate, while others reported "good" correlations with instrument measurement; (b) research assessing visual ranking of off-white fabrics with standard instrument measurement has not been located; (c) research testing the influence of different number of specimens in the ranking series has not been located; (d) instrumentation cannot directly test and totally replace human visual responses even though it is cheaper. This researcher attempted to bridge these gaps by investigating the accuracy of the ranking procedure to detect colour differences in off-white fabric specimens when panelists are allowed to manipulate the specimens.

Research Questions

The following three research questions were formulated for this study:

1. Does a change in the colour difference value of adjacent specimens affect the accuracy of the visual assessment in a ranking process?

2. Does a change in the number of specimens being ranked at one time affect the accuracy of the visual assessment in a ranking process?

3. Does a panelist's level of visual perception affect the accuracy of the visual assessment in a ranking process?

Hypotheses

In order to answer the research questions raised for this study, three hypotheses were proposed. They were:

Ho_1: There is no difference in the strength of relationship between visual and instrument ranking orders when the colour difference value of consecutive specimens changes.

Ha_1: The strength of relationship between visual and instrument rankings increases as the colour difference value of consecutive specimens increases.

Ho_2: There is no difference in the strength of relationship between visual and instrument ranking orders when the number in series of specimens being ranked varies.

Ha_2: The strength of relationship between visual and instrument rankings decreases as the number in series of specimens being ranked increases.

Ho_3: There is no difference in the strength of relationship between visual and instrument ranking orders for different levels of subject's visual perception.

Ha_3: The strength of relationship between visual and instrument rankings increases as the levels of subject's visual perception increases.

Definitions

Only the variables or concepts under investigation will be defined.

1. Near white is any colour having a Munsell Value greater than 8.3 (luminous reflectance 65%) and Munsell Chroma no greater than 0.5 for B hues, 0.8 for Y hues, and 0.3 for all other hues (ASTM E 313-87).

2. Off-white is a series of colours having a Munsell Value greater than 8.0 (luminous reflectance 62.4%) and Munsell Chroma no greater than 0.5 for 5GY hues. (The range of Yxy colour space values for the study are: Y = 62.40 to 80.30, x = 0.3120 to 0.3138, y = 0.3176 to 0.3209.)

3. Tristimulus values are the specified stimuli required to match a colour. In the CIE system, they are assigned the symbols X, Y, and Z (ASTM E 284-91).

4. Colour difference is indicated when either an instrument or a subject detects a change in a specimen's hue, chroma and/or lightness, the instrument value used in this study is ΔE representing the total colour difference.

5. Level of visual perception is the measurement of a subject's ability to detect colour differences correctly in the near white colour space.

6. Ranking method is where a series of three or more specimens are arranged into a deliberate order according to one specified characteristic. In this study, the characteristic is whiteness, and specimens are arranged from the whitest to the least white.

7. Paired comparison method determines if a colour difference can be perceived between two specimens. If a difference is noted, the subject identifies which specimen is the whiter of the two.

8. Ranking accuracy in this study is measured by the strength of relationship between a subject's ranking order and an instrument ranking for each series, calculated by Kendall's rank correlation coefficient, τ (tau).

Limitations of the Study

1. The results of this study may be influenced by the specimen mounting system, related to off-white colours only and associated with an artificial daylight source.

2. The results may not be generalized beyond those female student subjects who have taken part in the sensory evaluation.

Assumptions for the Study

1. The Hunterlab measurement can be used to determine the expected ranking of specimens from whitest to least white.

2. The CIELAB ΔE value is an appropriate measure to assess colour change in the off-white colour space.

3. The instrument value, CIELAB ΔE, is more sensitive to colour change than the human eye.

4. The off-white specimens produced for this study simulate the hues and chroma anticipated in an experimental series of detergency evaluations.

5. Subjects will use a form of paired comparison when they are allowed to manipulate the specimens into a rank order from the whitest to the least white.

6. Subjects favour a bluish hue in white fabric.

7. The physical and mental condition of observer, observing situation of the testing environment, geographical location, and fashion did not influence the results of the visual sensory evaluation on textile fabrics.

CHAPTER 2

LITERATURE REVIEW

This chapter includes two major sections: colour measurement theory, and applications to textile fabric whiteness. Only the fundamental theory of instrument colour measurement and visual evaluation used in this study are discussed. The applications cover only the located assessments of fabric whiteness.

Fundamental Theory of Colour Measurement

Theory of colour measurement is reviewed. The theory is applied to colour systems that relate instrument measurement to visual assessment. Factors which influence instrument measurement and visual evaluation are discussed.

Instrument Colour Measurement

This section describes trichromatic colorimetry, L, a, b colour space, colour difference formulae, and factors influencing instrument colour measurement.

Trichromatic Colorimetry

In 1931 the Commission Internationale de l'Eclairage (CIE) introduced the concepts of standard light sources, standard observer and method to generate instrument data. Prior to 1931, the CIE adopted the average r, g, b values, called the tristimulus values representing red, green and blue primaries respectively, for a small number of observers, as the experimental definition of the CIE 1931 Standard Observer. The term "observer" here refers to instrument measurement (Billmeyer & Saltzman, 1981). Experimentation confirmed that three primaries were enough to match or to identify any colour using both positive and negative values of the primaries.

To eliminate negative numbers within the tristimulus values, the "imaginary" primaries X, Y, and Z were identified as the 1931 CIE Standard Observer. This is a mathematical transformation from the original red, green, and blue primaries which no light source can produce. When these new tristimulus values specify a textile fabric colour, the maximum value assigned to Y is 100 which represents a perfectly reflecting white. However, the other tristimulus values, X and Z, can be larger than 100. One would find that X exceeds 100 for a perfect white with a yellowish hue and Z exceeds 100 for a perfect white with a bluish hue.

The CIE system plots colour on a chromaticity diagram with coordinates x and y which are defined as:

$$x = X/(X+Y+Z) \qquad y = Y/(X+Y+Z).$$

The third dimension, Y, of the system is added by an axis rising from the illuminant point of the chromaticity diagram to form a three-dimensional CIE system (Y, x, y). The system is used to match colour; two colours match only when they have exactly the same tristimulus values (Billmeyer & Saltzman, 1981). The CIE system does not describe hue, chroma and value of a colour, but remains the basis of modern colorimetry. ASTM Designation E 313 (1987) defined the range of near white using the CIE system. Off-white is also defined according to the CIE system.

L, a, b Colour Space

Hunter derived an opponent colour coordinate (L, a, b) measurement system where the "L" axis measures the specimen's lightness and varies from 100 for perfect white to zero for black, "a" measures the red-green complementary hues, and "b" measures the yellow-blue complementary hues (Hunter, 1975). This was a nonlinear transformation based on the CIE 1931 tristimulus values using the CIE standard illuminant C. Hunter's L, a, b system is probably the most widely used of all colour spaces except for the 1931 CIE Y, x, y system (Billmeyer & Saltzman, 1981).

The 1976 CIE L* a* b* colour space, a cube-root transformation to CIE 1931 tristimulus values, was officially recommended by CIE in 1976 and has been used mainly in the paint, plastic, paper and textile industries (Robertson, 1986). The CIELAB colour measurement system calculated the L* a* b* opponent-colour scales by applying a cube-root transformation to the CIE 1931 tristimulus values (ASTM E 284-91). These L* a* b* values describe a colour space related to lightness, redness-greenness and yellowness-blueness, respectively, of a specimen.

Colour Difference Formulae

Colour differences including both colour change and staining can be determined by instrument measurement or calculation (Stearns, 1974). Many colour difference formulae have been proposed over the years. Hoban (1981) indicated that the correlations between calculated values and visual colour differences were between 0.6 and 0.7 for the "better formulae." The 1976 CIELAB equations fall within this correlation range, and they were recommended by standards organizations such as ASTM (1989) and AATCC (1985) to promote consistency in colour communication.

Factors Influencing Instrument Colour Measurement

The precision of instrument colour measurement is affected by many factors. These are:

(a) differences between instruments (Hunter, 1961);

(b) reliability of the instrument (Harold, 1987);

(c) cleanliness of optical and electronic systems (Clemson Conference, 1990);

(d) accuracy of calibration (Hoban, 1981; Clemson Conference, 1990);

(e) electronic instability (Hoban, 1981);

(f) specimen presentation techniques (Clemson Conference, 1990);

(g) structure of fabric (Hunter, 1961; Stearns, 1974; Billmeyer & Saltzman, 1981; Clemson Conference, 1990).

Hoban (1981) and Harold (cited in Clemson Conference, 1990) both mentioned that inaccurate calibrations will lead to software errors. Hoban (1981) showed that instrument drift existed in the colour measurement due to electronic instability. This drift could not be corrected by data manipulation, and researchers must accept wider tolerances than they would desire. More recent research determined that re-calibration of modern solid-state instrumentation needs to be carried out only once or perhaps twice a day. This conclusion was confirmed by analyzing data continuously throughout the day and by demonstrating that the calibration L* values from Munsell value cards remained almost constant during the experiment (Palmetto Section, 1988).

In addition to colour, the texture, gloss and transparency of fabrics affect colour measurement. According to Stearns (1974), the observed colour was influenced by texture because of the variations produced by the geometry of illumination and viewing. Gloss influenced observed colour because light reflected from a coloured surface was different from that reflected from the dyes within the fibres. Finally, observed colour of transparent specimens was affected by the light reflected off the background colour.

Specimen presentation and variations within a presentation technique may create errors in colour measurement; written instructions could minimize variability in instrument or visual colour measurement (Clemson Conference, 1990). Hunter (1961), Matthews (1968), Stearns (1974), Billmeyer and Saltzman (1981) and Harold (1987) emphasized that specimen selection, physical dimensions, specimen preparation, specimen mounting, and specimen orientation should be closely controlled. The type of instrument illumination, viewing condition, and operator technique have also influenced the colorimetric results.

Test specimens should be opaque to the instrument detector which could require multiple layers of fabric (AATCC Test Method 153-85). The instrument assessment should distinguish no change in reflectance if fabric layers were increased. The specimen dimensions must cover the measuring port or the measuring beam.

In general, textile materials are not opaque. Black backing panels may lower the instrument reflectance values whereas white backings may increase the reflectance values. The black backing surface absorbs the light transmitted through the specimen, while the white one reflects light back through the specimen. For visual assessment, the standard method (ASTM D 1729-89) suggested that the backing and surrounding area should have a neutral colour with a Munsell chroma value less than 0.3. This method did not recommend any procedure (e.g., increasing the number of layers) to overcome the backing influence.

Matthews (1968) proposed that each specimen should be measured in four different areas, and that the specimen and reference should be protected in a dark place when they are not being tested. AATCC (Test Method 153-85) specified that at least two measurements should be made on each specimen; whereas, ASTM (D 2244-89) recommended a minimum of three test areas on each specimen. Hoban (1981) reported that some fabrics were highly dependent on their testing orientation and suggested that colour measurement of a given specimen area could be improved by averaging readings which have been recorded at 90 degrees to one another.

Visual Evaluation

Visual evaluation is one kind of sensory evaluation. When conducting sensory evaluation, the researcher assesses some factor effect, which could be a measurement approach (e.g., the number of specimens being evaluated) or an ingredient change (e.g., the degree of colour difference). Methods such as paired comparison, triangle test, scaling and ranking, can be used to solve different sensory problems. The most commonly used methods for visual whiteness evaluation in textiles are ranking and paired comparisons.

Ranking Method

When ranking, a series of three or more specimens are arranged in a deliberate order according to one specific characteristic. Ranking is rapid and can be used to test several specimens at one time. Specimens are evaluated only in relation to each other (Larmond, 1977).

A series of specimens should be presented in a random order, in which the attribute being tested is present in different levels and covers the range of interest. Rank orders cannot be used directly as a measure of intensity; however, they are amenable to significance tests. A minimum of eight panelists should be used with ranking procedures in sensory evaluations. If 16 or more were used, discrimination of the attribute under consideration would be improved (Meilgaard et al., 1987).

Several researchers applied the ranking method to assess fabric whiteness. Furry et al. (1961) compared visual ranking results to "b" opponent colour data and reported a "good" agreement between visual rankings and instrument readings. Rhode Island Section (1966) also reported quite a "good" agreement between visual and instrument rankings when ranking 10 specimens, but did not provide numerical data to support what is a "good" agreement. Rank-order correlation coefficients ranging from 0.56 to 1.00 were reported by Warfield and Hardin (1981); three of the four panelists reported correct responses for each of the rankings when 10 specimens were used. The ranking devised by Pelton (1989) showed statistically significant agreement among panelists' ranking of eight specimens, but no instrument data was reported in the research.

Paired Comparisons

The paired comparison test is one of the most frequently used attribute difference tests in sensory evaluation. In this test, panelists identify which of the two specimens has a more intensive attribute or which specimen is preferred. The specimen pairs can be presented simultaneously or sequentially to the panelists. Randomization should be considered in both the simultaneous and sequential approaches. The number of correct responses can be counted in the direction of interest. For instance, the percentage of correct responses for a whiter specimen can be counted and then analyzed statistically. Research reported that 15 or more panelists are enough for statistical analyses (Galbraith et al., 1987; Meilgaard et al., 1987).

A paired comparison method was used by Coppock (1965). The visual evaluation illustrated "excellent" agreement with the measured whiteness values. Morris (1970) found correlations between rank orders generated by visual paired comparisons and instrument readings in the range of $\hat{\tau} = 0.58$ to $\hat{\tau} = 0.89$ when comparing visual evaluation to Rd measurements. When visual evaluation was compared to b values, $\hat{\tau}$ ranged from -0.55 to -0.89. White, Prato, and Morris (1984) did not report any correlations of visual and instrument assessment. Galbraith et al. (1987) correlated visual paired comparison on whiteness of 39 specimens to ΔE readings and reported correlations above -0.90 when ΔE varied from 0.4 to 17.6. Kim, Smith, and Spivak (1987) summarized visual paired comparison results and instrument measurements. However, no correlation between the visual and instrument values was calculated.

Factors Influencing Visual Colour Evaluation

Several factors may influence the results and interpretations of visual colour evaluation. These are:

(a) the gloss and texture of a surface (Stearns, 1974);

(b) degree of surface contamination (Meilgaard et al., 1987),);

(c) the relative sizes of contrasting colour areas (Burnham, Onley, and Witzel, 1970);

(d) the influences from adjoining or background colour (Billmeyer & Saltzman, 1981);

(e) the consistency of responses from panelists (Meilgaard et al., 1987);

(f) colour difference thresholds among panelists (Stearns, 1974; Meilgaard et al., 1987);

(g) colour vision deficiencies (Meilgaard et al., 1987);

(h) visual evaluation process (O'Mahony, 1986; Meilgaard et al., 1987).

In visual assessments, Burnham et al. (1970) found that placing specimens closer together magnified the contrast effect and the resulting colour difference. Galbraith et al. (1987) stated that consumers were more sensitive to increases in yellow discolouration of white fabrics than they were to greying.

Coppock (1965) reported gender difference in colour perception and illustrated that women were more precise in their evaluations, and more aware of whiteness differences than men. Warfield and Hardin (1981) indicated that visual perception varied from panelist to panelist.

Meilgaard et al. (1987) suggested that a prescreening of panelists would improve a panel's ability to detect small differences in the attribute. Panelists should receive written instructions before commencing the tests so that they would understand the procedures and could recognize the important differences of the attribute under investigation.

Keesee and Harold (cited in Clemson Conference, 1990) indicated that temperature, moisture and aging of textiles could influence the results in colour measurement. Coppock (1965) suggested that observer, observing situation, geographical location, and fashion were variables which could influence the results of the visual sensory evaluation on textile fabrics.

According to Meilgaard et al. (1987), one should control the handling, preparation and presentation of each specimen so that the procedure did not introduce extraneous effects and conceal unknown factor effects. One also might need to examine all kinds of apparent unrelated influences, including physical and mental conditions of the panelists, and the condition of the testing environment before commencing the assessment. Sensory evaluation performed by large panels has been valuable in predicting consumer reactions (Larmond, 1977).

Mental fatigue has limited the number of multiple sensory assessments attempted at one time. To avoid this fatigue, Coppock (1965) suggested that each evaluation session should be limited to 15 minutes or less.

Applications to Textile Fabric Whiteness

Colour difference can be measured by both instrument and visual sensory techniques. Regardless of the technique used, colour difference measurement can be divided into examination and assessment phases (Billmeyer & Saltzman, 1981). The examination phase requires a source of light, a standard versus the objective being measured, and a detector (visual or instrument). The assessment of colour differences requires the researcher to decide whether a difference exists, to describe the difference, or to determine whether the difference is acceptable.

In textiles, a number of researchers have investigated fabric whiteness. Some have compared instrument measurements with visual evaluations as illustrated in Table 1. Table 1 also summarizes the types of instrument readings and sensory evaluation techniques used to measure colour, colour differences and colour changes in white textile fabrics over the past 30 years. Researchers (Coppock, 1965; Galbraith et al., 1987; Morris, 1970; Rhode Island Section, 1966; Warfield & Hardin, 1981) have correlated visual evaluation with instrument measurement and the comparisons were reported as "satisfactory" in the studies. Table 1 demonstrates that the total colour difference value ΔE was used; but the values were calculated by different formulae. For instance, Rhode Island Section (1966) compared MacAdam ΔE to ranking results and Galbraith et al. (1987) compared CIELAB ΔE to paired comparison.

A summary of instrument and visual assessments is found in Appendix 1. The majority of researchers did not indicate the number of specimen layers which they had used. For visual evaluation, different light sources and viewing conditions were applied. Different layers and physical dimensions of specimens were viewed by different numbers of trained or untrained panelists. This review of literature showed little consistency in experimental procedures used by different researchers.

Table 1
Summary of Selected Research on Whiteness Measurement

Researcher	Topic	Instrument readings	Visual assessment
Furry, Bensing & Johnson (1961)	Colour effect in near white fabric	Hunter Rd, a & b	Ranking
Coppock (1965)	Whiteness scale Brightness	Purity	Paired comparison
Rhode Island Section (1966)	Whiteness of fluorescent brighteners	MacAdam unit ΔC, ΔL & ΔE	Ranking
Morris (1970)	Laundering cotton fabrics	Rd & b	Paired comparison
Carver & Wylie (1980)	Fabric whiteness	L, a, b,ΔE & W	
Warfield & Hardin (1981)	Fabric whiteness	Rd	Ranking
Morris & Prato (1982)	Soil removal	CIELAB ΔE & YI	
Paek (1983)	Soil removal	Reflectance Change	
White, Prato & Morris (1984)	Home laundering	Rd, ΔE (CIE), WI & YI	Paired comparison
Morris & Prato(1985)	Soil removal	Rd, a, b & WI	
Wilcock & Van Delden (1985)	Laundering performance	L, a & b	
Galbraith, Swartzlander & Hardin (1987)	Perception of colour changes	CIELAB ΔE & FMC 2ΔE	Paired comparison
Kim, Smith & Spivak (1987)	Comparative study on detergents	L, a, b, Y & WI	Paired comparison
Wilson (1987)	Fabric whiteness		Paired comparison
Lovingood, Woodard & Leech (1989)	Performance of laundry detergents	ΔE, L, a & b	
Pelton (1989)	Laundry detergent effectiveness		Ranking
Brown, Cameron, Meyer & Umber (1991)	Washing Quality	WI	

CHAPTER 3

METHOD

This chapter describes the procedures performed for the preliminary work, specimen preparation, selection of subjects for visual evaluation, and the experimental design used in the research. A summary of the activities is provided at the end of the chapter. The Statistical Analysis System (SAS) was used to analyze the data.

Preliminary Work

Before the research questions could be addressed, several procedures needed to be pretested in establishing the reliability (drift, precision and accuracy) of the instrument readings. The instrument's ability to detect minute colour differences in whiteness was also identified. A computer program was developed to replace the conventional method of collecting colorimetric instrument readings. The computer program minimized data collection time, facilitated data handling, reduced specimen contamination and minimized error resulting from the light source aging.

Instrument Reliability

The study required a series of reliable instrument reference measurements from the Hunterlab colorimeter (Model D25M - 9). According to the Hunterlab manual, drift and precision are determined by measuring the deviations from the standard white tile. The standard for drift must be less than ± 0.1 Hunter L, a, or b units in one hour or ± 0.3 in three hours. Precision must have standard deviations of 0.1 or less for the Hunter L, a, and b readings. The accuracy calculation of the instrument involving all six standard tiles must be less than 0.7 root mean square deviation units (Hunterlab Manual). Preliminary tests showed that the Hunterlab colorimeter met these specifications. Details of the tests are included in Appendix B.

No information about the reliability of CIELAB ΔE measurement could be located. Because the threshold test and the visual ranking evaluation required the CIELAB ΔE measurement, the researcher assumed that if the Hunter L, a and b values satisfied the reliability specifications, then the CIELAB ΔE measurement (Table B-1) would also be reliable because Hunter L, a, b and CIELAB ΔE are calculated from tristimulus values X, Y, and Z.

Computer Program for Instrument Data

A computer program (Appendix C) was derived from the AATCC Test Method 173-89 to calculate all necessary colour space values from the CIE tristimulus values X, Y, and Z. The values computed from X, Y, and Z included L, a, b, L*, a*, b*, ΔE (CIE), ΔL (CIE), ΔH (CIE), ΔC (CIE), x, and y. The researcher checked the accuracy of these values by comparing them to those recorded from the Hunterlab instrument for each of the six standard tiles (Table C-1).

Each value had a specific purpose. Hunter L, a, and b values determined the instrument reliability throughout the specimen measurement. CIE Y, x, and y values verified that the dyed specimens fell within the defined off-white colour space. CIELAB L*, a* and b* values assisted in the selection of specimens for the threshold and ranking procedures. The CIELAB ΔE values identified the colour differences between consecutive specimens, which were used later to select specimens for the visual threshold and ranking assessments. The differences in lightness, chroma, and hue were assessed by ΔL, ΔC, and ΔH to meet the criteria of specimen selection for the ranking tests.

Specimen Preparation for Visual Evaluation

This section describes specimen development which includes dyeing, mounting, measurement and selection of specimens for both the threshold and ranking experiments. All specimens were taken from a bleached mercerized cotton fabric, Type W405, supplied by Testfabrics, Middlesex, New Jersey. The fabric structural specifications were:

Fibre content: 100% cotton

Fabric structure: plain weave

Fabric count: warp: 20.3/cm, weft: 19.8/cm

Mass: 156.6 g/m^2

Specimens were cut into rectangular shapes with the long side representing the warp direction. The bevelled-cut in the top right corner identified the designated face of the fabric. Raw edges were overlocked to prevent yarns from fraying during wet processing. All specimens were prewashed in an Atlas launder-ometer with distilled water, using the temperature and time required for the dyeing procedure.

Specimen Dyeing

The goal of the dyeing experiment was to produce a series of fabrics varying in their degree of whiteness but having the same hue. These fabrics were then used to investigate the accuracy of the ranking procedure. Controlling colour difference (ΔE) in whiteness by dyeing meant that all the colour components ΔL, ΔC, and ΔH were controlled. However, the researcher wanted a smaller tolerance level in the hue than in lightness and chroma because the human eye had higher sensitivity to changes in hue (Harold, 1987). If the CIELAB co-ordinates a*, and b* for a dyeing series form a straight line and this line passes through the origin, the specimen series has the same hue (Broadbent, personal communication, June, 1992). Hue could be controlled by manipulating the dyestuff mixture.

Mixtures of red, yellow and blue reactive dyestuffs at various concentrations created a series of specimens which varied from a yellowish white to a bluish white. Each concentration was diluted ten times to create a specimen series in which each consecutive specimen had an increased ΔE from the undyed reference specimen. All dyeings were carried out in the Atlas launder-ometer using procedures recommended for reactive dyestuffs. Details of the dyeing are included in Appendix D. This dyeing technique generated off-white fabric specimens similar to those reported in Warfield and Hardin (1981) in which colour differences between specimens were controlled.

Specimen Mounting

To eliminate the backing influences for both instrument measurement and visual evaluation, a procedure similar to that outlined in AATCC Test Method 153-85 was followed. Specimens had to be mounted in such a way that they were opaque; otherwise, light would be transmitted through textile fabrics and reflect off either the instrument backing tile or the viewing cabinet surface. The texture and colour of these two backing surfaces were different and could influence the results (Cutler & Davis, 1972).

The washed, undyed standard fabric (W405) was used for this exercise. A designated specimen was always adjacent to the instrument orifice throughout all the measurements. Specimens were added one layer at a time behind the designated specimen up to 10 layers. The ΔE readings as each layer was added were generated first with the standard white tile as backing and then with the standard black tile as backing. The researcher stopped after the tenth layer because the ΔE values were very similar regardless of the colour of the backing tile.

The reference ΔE was the mean value of 10 layers of specimens backed by the white tile; it was given a value of zero. The ΔE values for the other layers were recorded in Appendix E and compared to the reference reading. The plot of ΔE versus the number of layers (Figure 1) illustrated the influence of the white and black tile backings on ΔE readings. Figure 1 showed that the backing influence was apparent up to nine layers. However, backing the specimens with nine layers of fabric was impractical because subjects would have to interchange a thick pile of specimens.

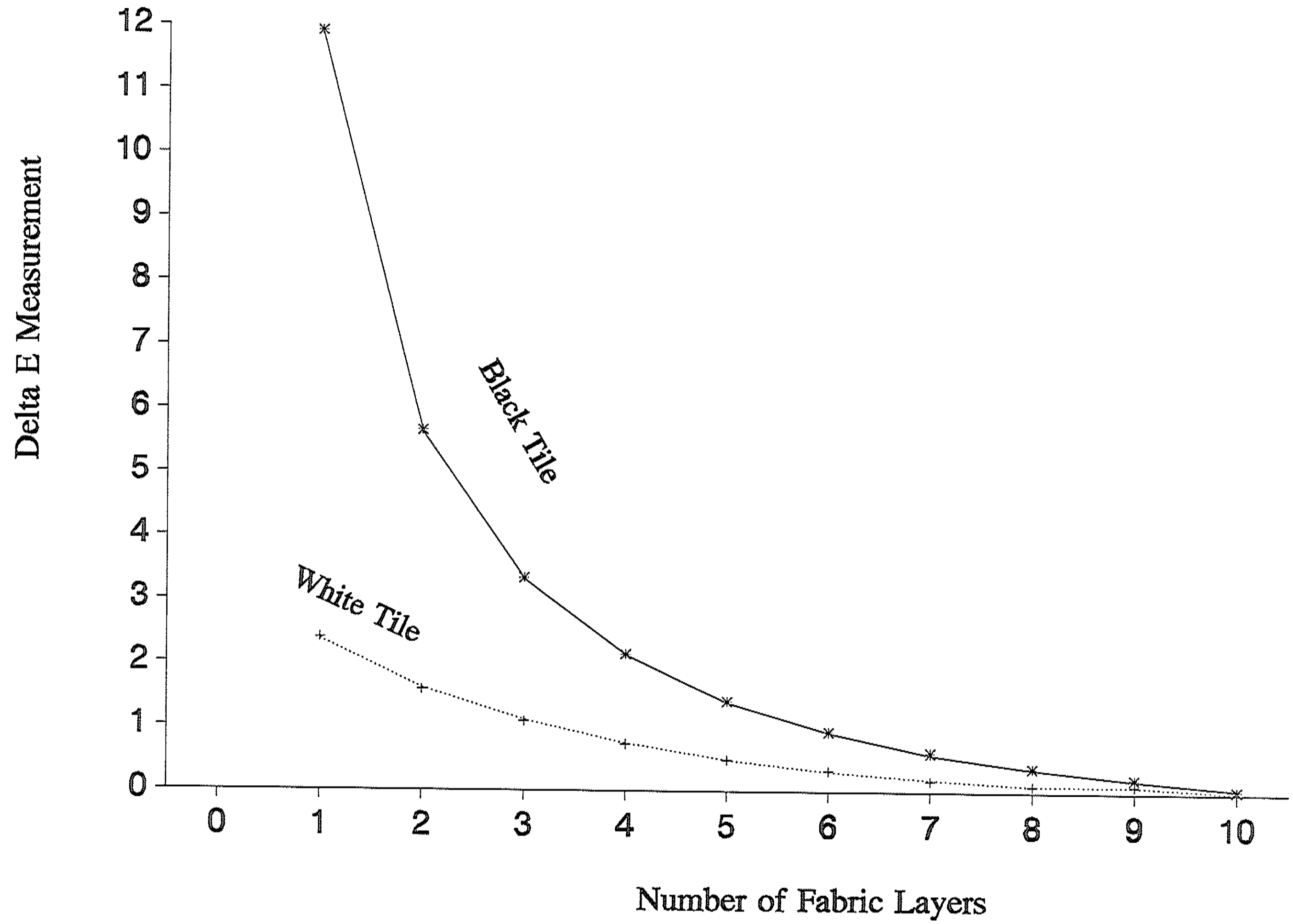

Figure 1. Influence of the number of layers and backing colour on ΔE

Also, in the given time frame, the researcher would not be able to produce nine or more replicas with the same CIE X, Y, and Z readings for each specimen. As an alternative, the researcher adapted the mounting system reported in Vanderhoeven's study (1992). Specimens were mounted on thick rigid cardboard using double-sided carpet tapes. Each specimen cardboard system was cut to the dimension (165 mm by 90 mm) recommended for visual assessment (ASTM D 1729, 1989). This cardboard mounting system provided the same opaque specimen for both instrument measurement and visual evaluation.

CIE Tristimulus Measurements of Specimens

After each series of dyeing, specimens were immediately mounted and measured. X, Y, and Z measurements were taken at three different places on each mounted specimen, first in the warp and then in the filling direction. During these measurements, the instrument reliability was checked every 15 minutes. If the standard white tile L, a, and b values did not fall within the instrument specification for drift, the data would be discarded. The instrument then would be recalibrated and the specimens remeasured. After each set of measurement, specimens were put into plastic bags and stored in a dark, dry place.

The instrument values could be fitted by a model involving specimen and orientation effects, and interaction between specimen and orientation. Specifically, the model is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

where

Y_{ijk}: the ΔE of the kth area of the ith specimen and jth orientation.

μ: the grand mean.

α_i: the main effect of specimen i.

β_j: the main effect of orientation j.

$(\alpha\beta)_{ij}$: the interaction of specimen i x orientation j.

ε_{ijk}: the experimental error.

The researcher chose the colour difference measurement (ΔE) over the whiteness index (WI) because the whiteness index was limited to the near white colour space according to the AATCC Test Method 110 and ASTM E313-87. One would expect that the interaction between specimen and orientation would be insignificant, and that the specimen orientation and the experimental error would have an insignificant contribution to the total variance of the measurement.

Specimen Selection

The computer program (Appendix C) assisted in the selection of specimens. For this study, the total colour difference between two specimens was defined as ΔE, derived from the CIELAB opponent colour scales and calculated (ASTM D 2244-89) by:

$$\Delta E = [\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}]^{1/2} \qquad \text{(Equation 1)}$$

ΔE also represented differences in lightness (L), chroma (C) and hue (H) which were represented by:

$$\Delta E = [\Delta L^2 + \Delta C^2 + \Delta H^2]^{1/2} \qquad \text{(Equation 2)}$$

Instrument values identified colour differences within each specimen and between consecutive specimens. The criteria of specimen selection for the visual evaluation were: (a) each specimen was compared to one common reference specimen; (b) ΔE readings within each specimen should be in the range of ± 0.04 (instrument drift established in Table B-1) from the average ΔE; (c) the nominal ΔE between consecutive specimens should increase in equal increments from the whitest to the least white for both threshold and ranking evaluations; (d) ΔL, ΔC and ΔH between consecutive specimens were all decreasing from whitest to least white specimen for visual ranking assessment.

Selection of Subjects

Before approaching students, approval for the study was received from the Ethics Review Committee (Appendix F). Female students enrolled in the Faculty of Human Ecology at the University of Manitoba during the 1993 winter semester were invited by letter (Appendix F) to form the required panels. Female subjects were chosen because they tend to have more sensitive visual perception in whiteness than do males (Coppock, 1965).

The letter outlined the requirements and duties for subjects and the time commitment. Students who agreed to participate in the study returned the signed consent form (Appendix F) and made appointments with the researcher to conduct the first of five testing sessions.

Visual Evaluation Sessions

The visual evaluation consisted of five sessions. Table 2 summarizes the activities in each session.

Table 2

Summary of Visual Testing

Session	Activities	ΔE	Number of specimens
1	Pseudo-isochromatic Plates 3 threshold tests		
2	2 threshold tests Completed questionnaire Ranking tests (vertical)	ΔE_{min}	3, 6, 9, 12
3	Ranking tests (vertical)	ΔE_{med}, ΔE_{max}	3, 6, 9, 12
4	Ranking tests (horizontal)	ΔE_{min}, ΔE_{med}, ΔE_{max}	3, 12
5	Ranking tests (horizontal)	ΔE_{min}, ΔE_{med}, ΔE_{max}	6, 9

The first session screened out subjects who might have visual colour deficiencies by asking them to view a number of Pseudo-isochromatic Plates. The Pseudo-isochromatic Colour Test required the subjects to respond to 46 alpha or numerical motives embedded in different coloured patterns. The test assessed the presence of visual colour deficiencies in the red-green and/or the yellow-blue hues. Subjects were aware that they might participate in one or five testing sessions. Those who were screened out after the first session would not be informed of their visual colour deficiencies unless they asked the researcher. Only then would the researcher identify the reason.

Subjects who correctly identified all the characters in the Pseudo-isochromatic Plates proceeded to the next session. The subjects came to the laboratory at the appointed time to assess colour differences in the off-white colour space by paired comparison and ranking methods. Subjects completed a brief background information questionnaire (Appendix F) which the researcher later used to describe the characteristics of subjects. Although this subject selection technique produced a convenience sample, the students who agreed to participate were randomly assigned to their visual threshold and ranking tasks.

Total confidentiality of results was assured by the researcher. Data were coded for analyses and no identification of individual subjects was possible. Only the researcher had access to the subjects' names and coding sheets as they were stored in a private location. When the study was completed, all records of subjects' names, telephone numbers, and addresses were destroyed.

Experimental Design

This section describes the randomized complete block design used in the ranking test, visual ranking protocol, and summary of the method. The dependent variable in this study was the subject's visual rank order. The independent variables were the number of specimens within the ranking series and the ΔE between consecutive specimens. A separate experiment was done on the level of visual perception. The fabric source, the specimen mounting system, the working range in colour space, and the viewing condition were held constant.

Randomized Complete Block Design for Ranking

To investigate the effects of ΔE between consecutive specimens and the number of specimens on visual rank order, a randomized complete block design was used (Meilgaard et al., 1987). A randomized complete block design is of the form:

	Factor						
Block	1	2	3	...	j	...	k
1	X_{11}	X_{12}	X_{13}	...	X_{1j}	...	X_{1k}
2	X_{21}	X_{22}	X_{23}	...	X_{2j}	...	X_{2k}
3	X_{31}	X_{32}	X_{33}	...	X_{3j}	...	X_{3k}
	.						.
	.						.
	.						.
i	X_{i1}	X_{i2}	X_{i3}	...	X_{ij}	...	X_{ik}
	.						.
	.						.
	.						.
b	X_{b1}	X_{b2}	X_{b3}	...	X_{bj}	...	X_{bk}

In this study, subjects were the blocks. Each subject evaluated all the specimens by ranking. X_{ij} represented the relationship between subjects' rankings and expected (instrument) rankings, defined by Kendall's Tau value of the subject i under the given factor j.

Factors for the Experimental Design

The model proposed to address the first two research questions can be illustrated by the following diagram. The two factors in this model were ΔE between consecutive specimens and the number of specimens being ranked at one time.

The first factor depended on the range of ΔE values established by the subjects' visual threshold test (ΔE_{min}, ΔE_{med}, and ΔE_{max} respectively). The values of ΔE_{min}, ΔE_{med}, and ΔE_{max} were kept constant for every subject throughout the ranking

Number of Specimens	Colour Difference		
	ΔE_{min}	ΔE_{med}	ΔE_{max}
3			
6			
9			
12			

test. The researcher determined the number of specimens to be 3, 6, 9, or 12. Therefore, a treatment was one ΔE-number of specimens combination in each cell. The experiment was done two times, once "vertically", and then "horizontally".

Visual Threshold of Colour Difference

Visual threshold was defined as the smallest colour difference which the eye could detect in the near white colour space. The threshold test determined the ΔE

values to be used as a factor in the experimental design. The results were also used to assist in the analysis of the ranking process. The development of the threshold test involved a pretest and the final test.

Threshold pretest. The pretest was carried out over a one-week period with six subjects who were not part of the final panel. During the first exercise, subjects observed specimen pairs and were asked to identify the whiter of two specimens without moving them. In another exercise, they were asked to identify colour differences (yellowish or bluish white) between pairs of specimen and were encouraged to interchange the positions of the specimens on the viewing board.

When subjects perceived a colour difference in the first exercise, the researcher asked the subjects to identify which of the two was whiter to confirm that the subjects were not guessing in the pretest. However, the researcher observed that the subjects had difficulty in identifying which was whiter. When subjects were asked to determine which was the whiter of the two specimens, they could have been responding differently to changes in ΔL, ΔC, and ΔH. By comparing subjects' responses to ΔE readings, it was not surprising that many mismatches were observed because ΔE was a function of ΔL, ΔC, and ΔH. Hunter (1961) reported that people preferred yellowish or bluish white over pinkish or greenish ones. Hunter (1961) also suggested that bluish white was preferred over yellowish white. The specimens used in the threshold test are along yellowish-bluish continumn, which was established by CIELAB a* and b* coordinates. The question "which of the two specimens is whiter" allowed subjects to choose either a yellowish or a bluish specimen. The question "which one appears yellowish or bluish white" gave the same hue orientation as the dyed specimen used. Therefore, the revised question was more effective than the first one in identifying subjects' ability to detect a difference in threshold test which corresponded to the ΔE instrument readings.

Another purpose of the pretest was to determine the specimen handling technique. Rhode Island Section (1966) suggested that panelists be allowed to manipulate the specimens in a paired comparison test. However, the paired comparison method reported by a Canadian detergent manufacturer (B. Shantz, personal communication, 1991) did not allow panelists to touch the specimens. Research comparing the two specimen handling techniques has not been located.

Figure 2 showed that when subjects were asked to identify the whiter of the two specimens and were not allowed to interchange the position of the specimens, they could only reproduce consistent correct responses when ΔE was equal to or greater than 1.0 ± 0.04. However, subjects correctly and consistently identified colour differences greater than 0.4 ± 0.04 when they were allowed to interchange the position of specimen pairs. The researcher concluded that interchanging the specimen positions improved the accuracy of the visual threshold in the near white colour space. The researcher also noted that the ranking protocol required a similar interchange or manipulation of adjacent specimens.

<u>Final threshold test</u>. To avoid fatigue, the threshold test was repeated five times over two sessions. In the first session, subjects repeated the threshold test three times and twice again in the second session. Each time, the specimens were presented

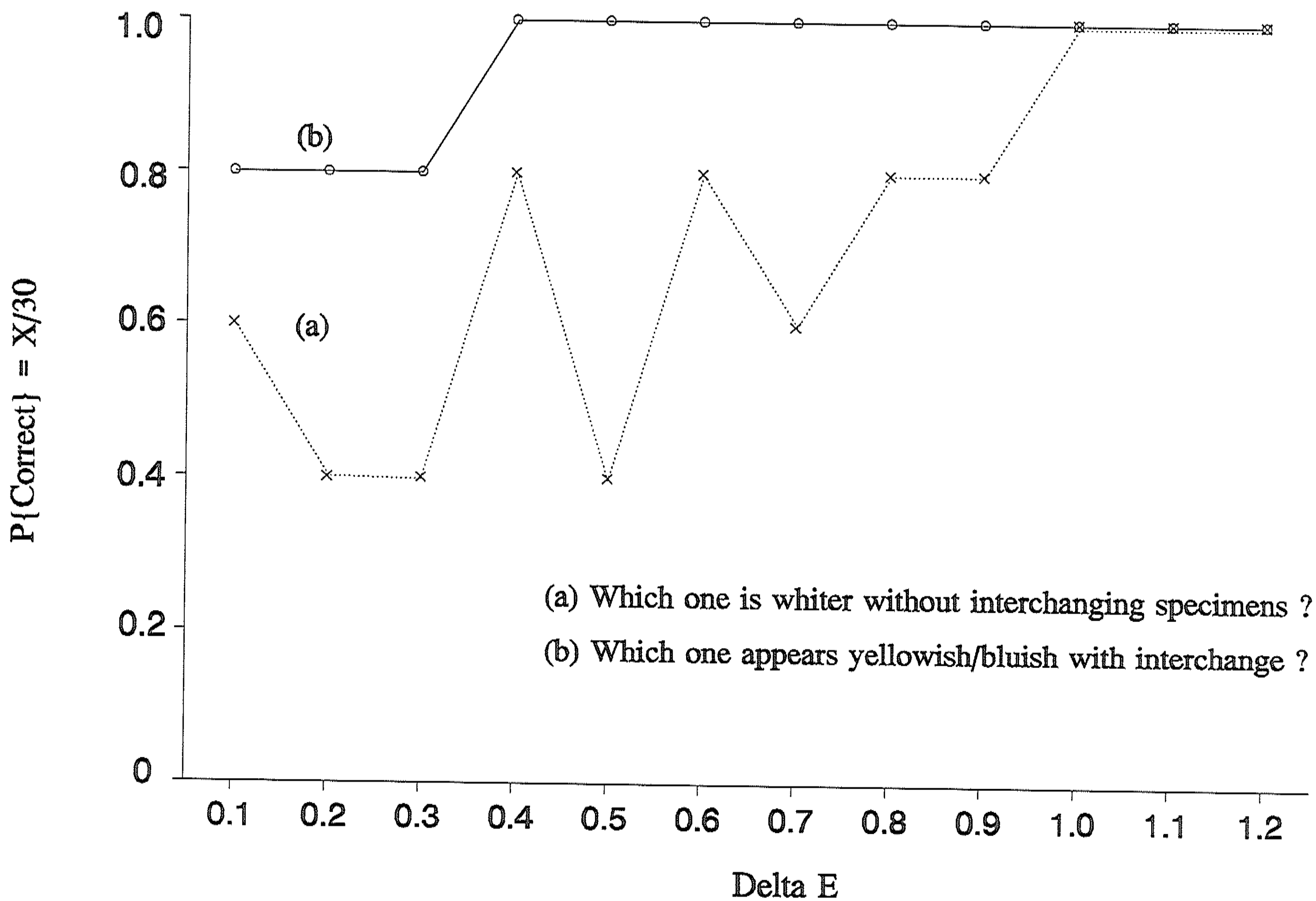

Figure 2. Preliminary threshold test - Possibility of correct response with increasing ΔE (five trials of each ΔE for six subjects)

in a random order.

In the visual threshold test, the paired comparison method was used. The subjects read the instructions (Appendix G) for the threshold test and were asked if they understood the directions before commencing the test. Specimen pairs were placed flat, side by side, on the viewing board under Macbeth. Subjects evaluated the colour differences of the specimen pairs at arm's length. The day-light-source illumination from the Macbeth was directly above the specimen area, and the viewing angle was about 45° from the perpendicular. Both the researcher and subject wore white lab coats to minimize light reflection from cloth and gloves to prevent contamination from their hands. All other lights in the laboratory were switched off.

Specimen pairs were presented to each subject independently in a random order. Subjects did not know that the series had a fixed reference specimen, which eliminated an expectation error (Meilgaard et al, 1987). Subjects were asked the question "Can you see a colour difference between these specimens?" When the answer was "yes", the subject was then asked "Which one appears yellowish or bluish white?" The researcher recorded the results as "✓" (meaning "correct") when the answer was the same as that from instrument measurement. If the panelist gave a different response from the instrument measurement, or remained undecided over 10 seconds, the result was recorded as "-" (no difference being identified) by the researcher.

The responses from each subject and from the panel were compared with the expected CIELAB ΔE values. The proportion of correct responses determined the

ΔE_{med} and ΔE_{max} values for the first factor. The derivation of ΔE_{min}, ΔE_{med}, and ΔE_{max} will be explained in the next chapter.

Visual Ranking Protocol

The visual ranking was done over four sessions, two for vertical data collection, and another two for horizontal data collection (Table 2). Before commencing the ranking test, each panelist read the instructions (Appendix G) and was asked if they understood their tasks. Specimen series were presented to panelists in different random orders. The specimens were placed flat side by side on the Macbeth viewing booth, and the rankings were made under the day-light-source illumination at arm's length by subjects. All other lights in the laboratory were turned off during the test. The illumination of day-light-source from the Macbeth was directly above the specimen area and the viewing angle was approximately 45° from the perpendicular. Illuminant C was used in both instrument and visual conditions as the light source. Both the researcher and the panelist wore white lab coats and gloves. The instrument rank orders were marked at the back of each specimen and were invisible to panelists.

Data Analysis

The visual rankings produced ordinal measurements that required nonparametric statistical analyses. To analyze the effects of ΔE between consecutive specimens and the number of specimens on visual rank order, Kendall's Tau correlation coefficient (τ) for ranking was calculated to determine the accuracy of a subject's ranking against an expected ranking (from instrument measurement) for each series. The ranking accuracy was defined previously under definitions. Page's test (Daniel, 1978) was used to test the first two hypotheses proposed for this study. Page's test would determine if a trend existed for the factor effects.

In this chapter, the researcher described the procedures for assessing reliability of the Hunterlab. The process of generating several colour space values that were critical for the visual ranking experiment was also described. Then, the researcher accounted for the steps in preparing the specimens, how the final specimens were selected, and how the subjects were recruited. The experimental design for the ranking tests was explained. Finally, Table 3 summarizes the chronological order for the experiments involved in this research.

Table 3

Chronology of Experiments

Sequence	Experiments	Sequence	Experiments
1	Instrument reliability	5	Selection of subjects
2	Computer programming	6	Visual threshold tests
3	Specimen dyeing & measurement	7	Specimen selection for rank
4	Specimen selection for threshold Test	8	Visual ranking tests

CHAPTER 4

RESULTS AND DISCUSSION

This chapter presents the results in four sections: specimen preparation, subjects, the threshold test, and the ranking test.

Specimen Preparation

This specimen preparation section reports the results from the specimen dyeing experiment, and CIE tristimulus measurements of dyed specimens.

Specimen Dyeing

The standard fabric (Type W405) was dyed using the procedure outlined in Appendix D with various mixtures of red, yellow, and blue reactive dyestuffs (Table 4) to create series of specimens which varied from the whitest to the least white. Each mixture used a dyestuff concentration of 0.032 g/1000 ml. The mixture proportion determined the hue. The procedure produced a series of ten dilutions (50:50) from the given dyestuff mixture to vary whiteness.

After the first series D-1 had been dyed, the CIELAB values L^*, a^*, b^* and ΔE were measured using a single specimen layer and the standard white tile as the backing. The a^* and b^* coordinates were then plotted on the graph (Figure 3). The a^* axis represented the red (+) and green (-) opponent colour scales, and b^*

Table 4

Selection of Dyestuff Mixture to Produce Specimens With Identical Hue

Dyeing code	Dyestuff proportion		
	Red	Blue	Yellow
D-1	1	1	1
D-2	1	1	2
D-3	1	0.5	3
D-4	1	1	3
D-5	0.5	1	3
D-6	0.5	1	2
D-7	0.25	1	2

axis represented the yellow (+) and blue (-) opponent colour scales. The dyestuff mixture for D-2 was based on the results from D-1 with the objective of producing specimen readings that pass through the origin. This dyeing exercise produced seven different specimen series which were plotted in Figure 3.

Figure 3 shows that the dyestuff proportions in Dyeing 3 and Dyeing 5 (corresponding to D-3 and D-5 in Table 4) produced erratic hue changes. The visual assessment of colour difference in these series did not coincide with ΔE measurements. The dyestuff proportion in D-6 (Table 4) produced the best specimen series because even though the series did not pass through the origin, but the straight line of Dyeing 6 was very close to the origin. The proximity of Dyeing 6 to the origin meant that the hue from dilution to dilution changed very slightly. The straight line indicated that both the hue and chroma changes were controlled, and hue moved gradually from

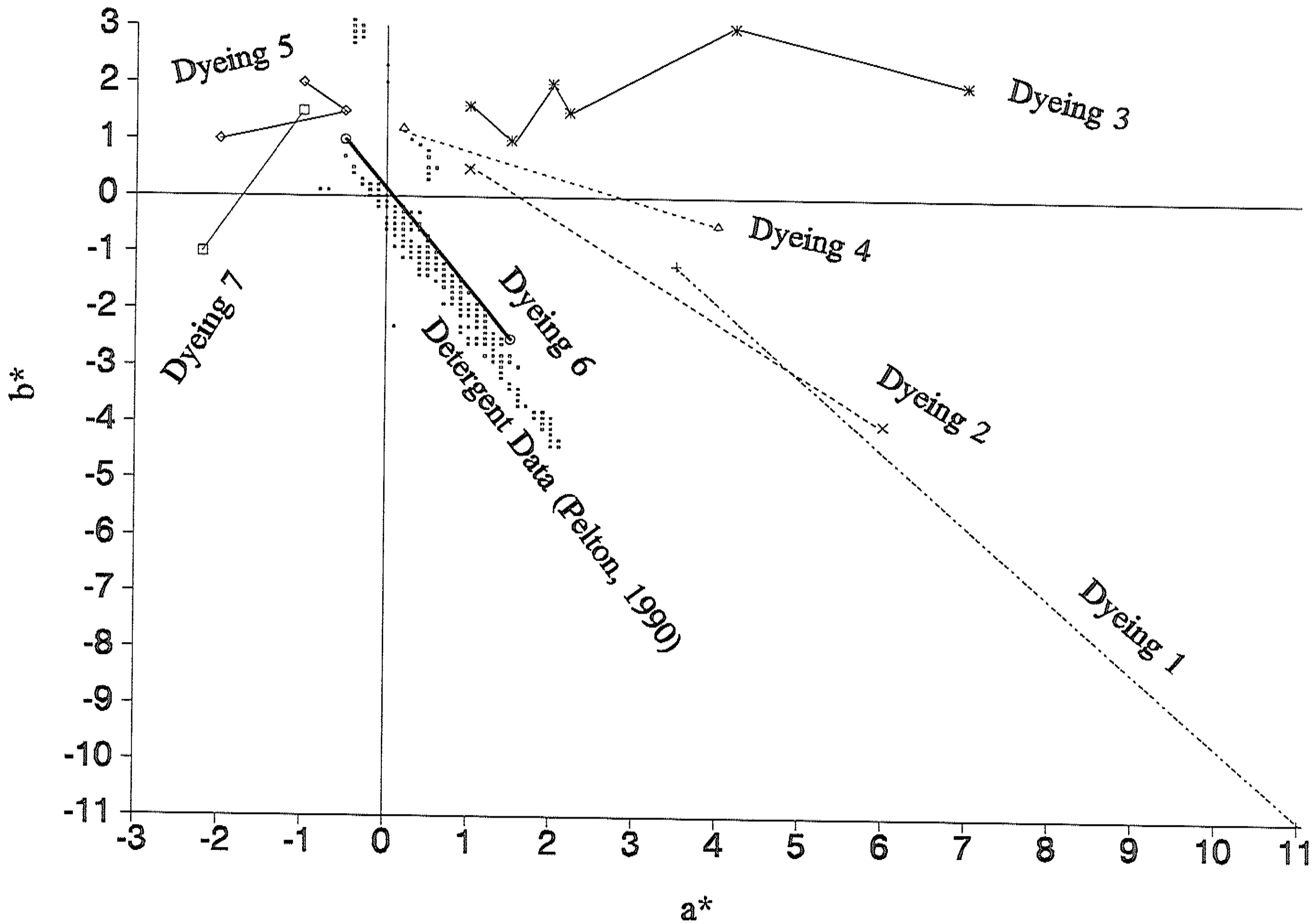

Figure 3. CIELAB a* and b* coordinates for dyeing pretest of controlling hue

Table 5

Dyestuff Concentrations and Dilutions Producing Off-White Specimens

	Dyestuff concentration g/ml		Dyestuff dilution	
Dyeing code	Dyestuff weight (g)	Starting volume (ml)	Dilution proportion(%)	Number of dilutions
S - 1	0.032	1000	25	10
S - 2	0.056	2000	22.5	10
S - 3	0.064	2000	25	10
S - 4	0.100	1400	25	10
S - 5	0.020	2000	18	10
S - 6	0.030	1600	21.9	10
S - 7	0.040	1600	21.9	10
S - 8	0.050	1600	21.9	10
S - 9	0.050	1600	21.9	10
S - 10	0.060	1600	21.9	12

yellowish to bluish (Hunter, 1965).

The researcher compared the a* and b* values generated from the above dyeings to those measured on a series of white fabrics which were subjected to different washing conditions (Pelton, 1990). Pelton's data produced a linear band from the second quadrant through to the fourth quadrant through the origin (Figure 3). Dyeing 6 produced a linear plot similar to this band. The dyestuff mixture from D-6 (Table 4) was then selected to generate off-white specimens for the visual sensory threshold and ranking experiments. The researcher manipulated the dyestuff concentration, dilution and the number of dilutions to produce specimens with varying ΔE values from the washed, undyed original fabric (W405). Table 5 listed the dyestuff concentration (g/ml), dilution proportion and the number of dilutions which were applied to the D-6 mixture (Table 4) using the dyeing procedure in Appendix D to create 102 off-white specimens.

CIE Tristimulus Measurements of Dyed Specimens

X, Y, and Z measurements were taken from three different places on each mounted specimen, first in the warp and then in the filling direction. The CIE tristimulus X, Y, and Z measurements for each dye series in Table 5 took approximately 40 minutes to complete. Instrument drift was checked periodically using the standard white tile procedures described in Appendix B. The computer program (Appendix C) generated the 13 instrument values L*, a*, b*, ΔE (CIE), ΔL (CIE), ΔC (CIE), ΔH (CIE), L, a, b, Y, x, and y for each measurements. In total, 7,956 values were produced (102 x 6 x 13 = 7956).

The CIELAB a* and b* values for the 102 mounted specimens suggested that the dyeing procedure controlled hue, and chroma. In Figure 4, the points scattered above and to the right of the origin indicated that the specimens were not of the same hue. However, when a* and b* values from Figure 4 were compared to those from D-6 in Figure 3, the dyeing procedure did control the hue, and chroma, and the mounted specimens had similar hues and chroma to the preliminary dyeing D-6.

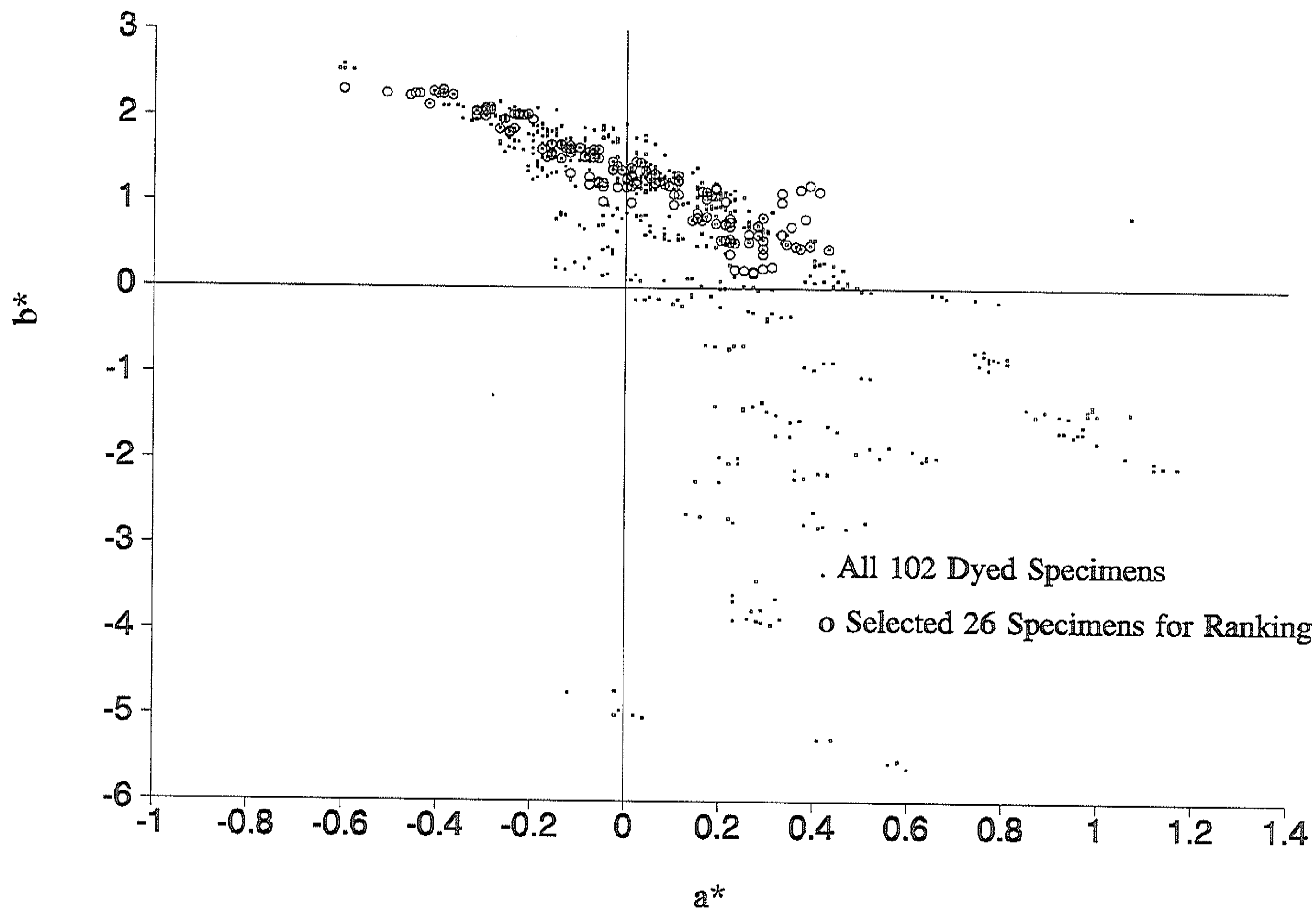

Figure 4. CIELAB a* and b* coordinates for mounted specimens

Subjects

One hundred twenty six letters of invitation were distributed to Human Ecology students. Thirty-five students agreed to participate in the study; nine students withdrew from the study at various stages. The final sample consisted of 24 subjects. Galbraith et al. (1987) suggested that a panel of 15 was large enough to produce a valid assessment in visual textile evaluations. Therefore, the data collected from 24 female subjects were assumed to be adequate for this study.

The subjects, from the Clothing and Textiles Department, ranged in age from 21 to 25. The majority had taken at least three textile science courses, and were responsible for doing their own laundry. All paid special attention to colour when purchasing clothes. Most of the subjects purchased predominantly white apparel items in the six months prior to participation in the study.

The Threshold Test

The purpose of the threshold test was to establish the ΔE values for the colour difference factor in the experimental design. Before the threshold test was carried out, the specimens had to be selected from the pool of 102 dyed specimens.

Specimen Selection

For the threshold test, the researcher was interested only in the total colour difference value ΔE. The following criteria were used for selecting specimens: (a) each specimen was from the near white colour space; (b) each was compared to one common reference specimen; (c) the variation of ΔE within each specimen was chose to be less than $\pm$ 0.04 from the calculated average; (d) difference in ΔE between consecutive specimens was 0.1.

From textile standards (AATCC Evaluation Procedure 1, 1987) and Hunterlab manual, one would anticipate that the smallest detectable colour difference in near white would be approximately 0.2 ΔE units. Preliminary threshold testing demonstrated that 1.0 ΔE unit might be the upper threshold limit for some student subjects to detect the colour differences between textile fabrics. The smallest detectable colour difference (4-5) for the grey scale to colour change corresponded to 0.8 $\pm$ 0.2 ΔE units giving a colour difference range from 0.6 to 1.0 ΔE units (AATCC Evaluation Procedure 1, 1987). Stearns (1974) and Crown (1978) also indicated that the threshold could vary from subject to subject.

Using the criteria above, and the threshold pretest results, twelve specimens were selected where the nominal ΔE ranged from 0.1 to 1.2 units. The reference specimen had instrument values of $L^* = 91.00$, $a^* = -0.04$, and $b^* = 1.84$. The computer program was used for specimen selection. Figure 5 showed the plot of the ΔE range for the 12 specimens. The ΔE ranges for consecutive specimens demonstrated no overlap (Appendix H).

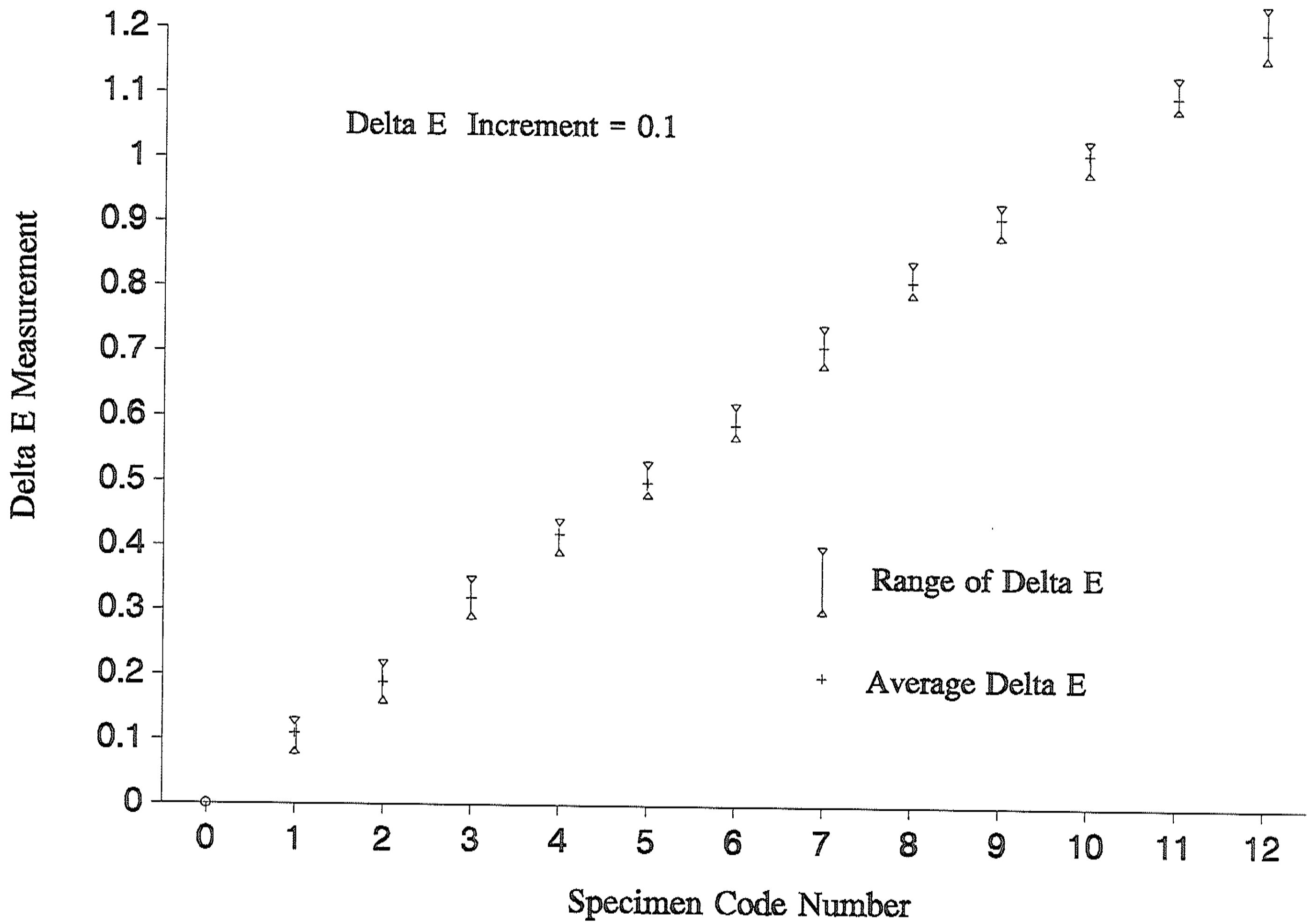

Figure 5. ΔE range of selected specimens for the threshold test

Visual Threshold of Colour Difference

When subjects were asked to detect colour differences between pairs of specimen, their responses were compared to the nominal ΔE values. Figure 6 illustrated the correct responses of the five comparisons for each nominal ΔE by 24 subjects in threshold test. The data are listed in Appendix H . No subjects had any error when the colour differences were greater than 0.5 ΔE units. Therefore, the researcher decided that ΔE_{min}, ΔE_{med}, and ΔE_{max} should centre around the 0.5 ΔE values. Consequently, the ΔE_{med} was assigned a value of 0.5. For ΔE_{min}, the value of 0.2 was assigned because according to the Hunter manual, 0.2 ΔE was the smallest detectable difference possible by the human eye. A value of 0.8 was assigned to ΔE_{max} because the researcher had a criterion of equal increments between each ΔE factor selected. The 0.8 ΔE also corresponded to the AATCC colour change grey scale 4-5 (AATCC Evaluation Procedure 1, 1987).

The researcher observed that percentage of correct response increased as ΔE increased. However, as Figure 6 has illustrated, subjects had difficulty identifying colour difference of 0.3 ΔE. To ascertain why the specimen with 0.3 ΔE caused problems, the researcher plotted ΔE, ΔL, ΔC, and ΔH of all 12 specimens. Figure 7 showed that as ΔE increased, ΔL, ΔC, and ΔH did not always decrease. In comparison to the reference specimen, a decreasing trend for ΔL meant that the specimens were getting darker; a decreasing ΔC signified that the specimens were getting duller, and a decreasing ΔH meant that the specimens were getting bluer. A closer look at the colorimetric measurements of the specimen with 0.3 ΔE revealed that its lightness and

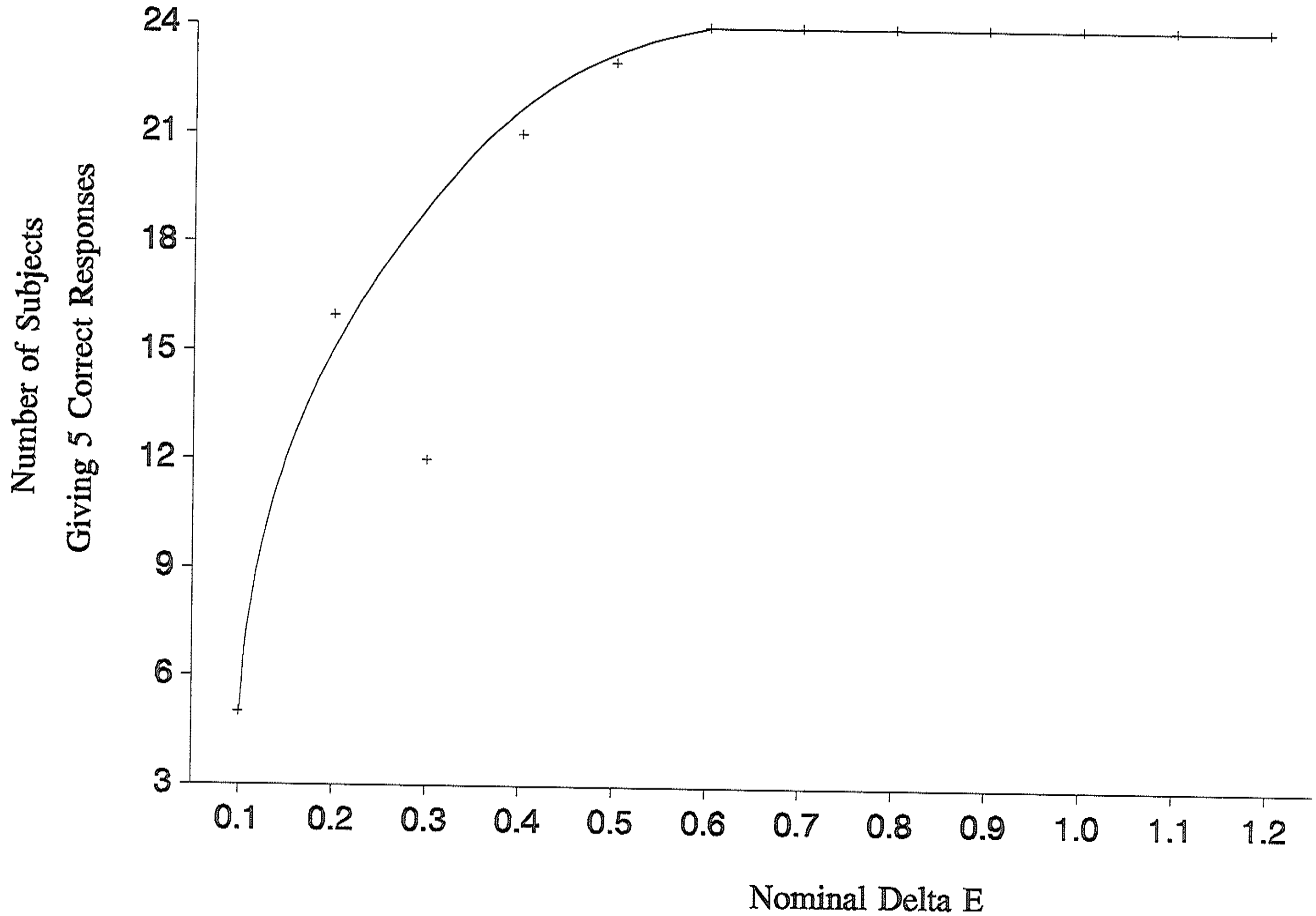

Figure 6. Distribution of correct responses per subject for each nominal ΔE of the final threshold test by 24 subjects.

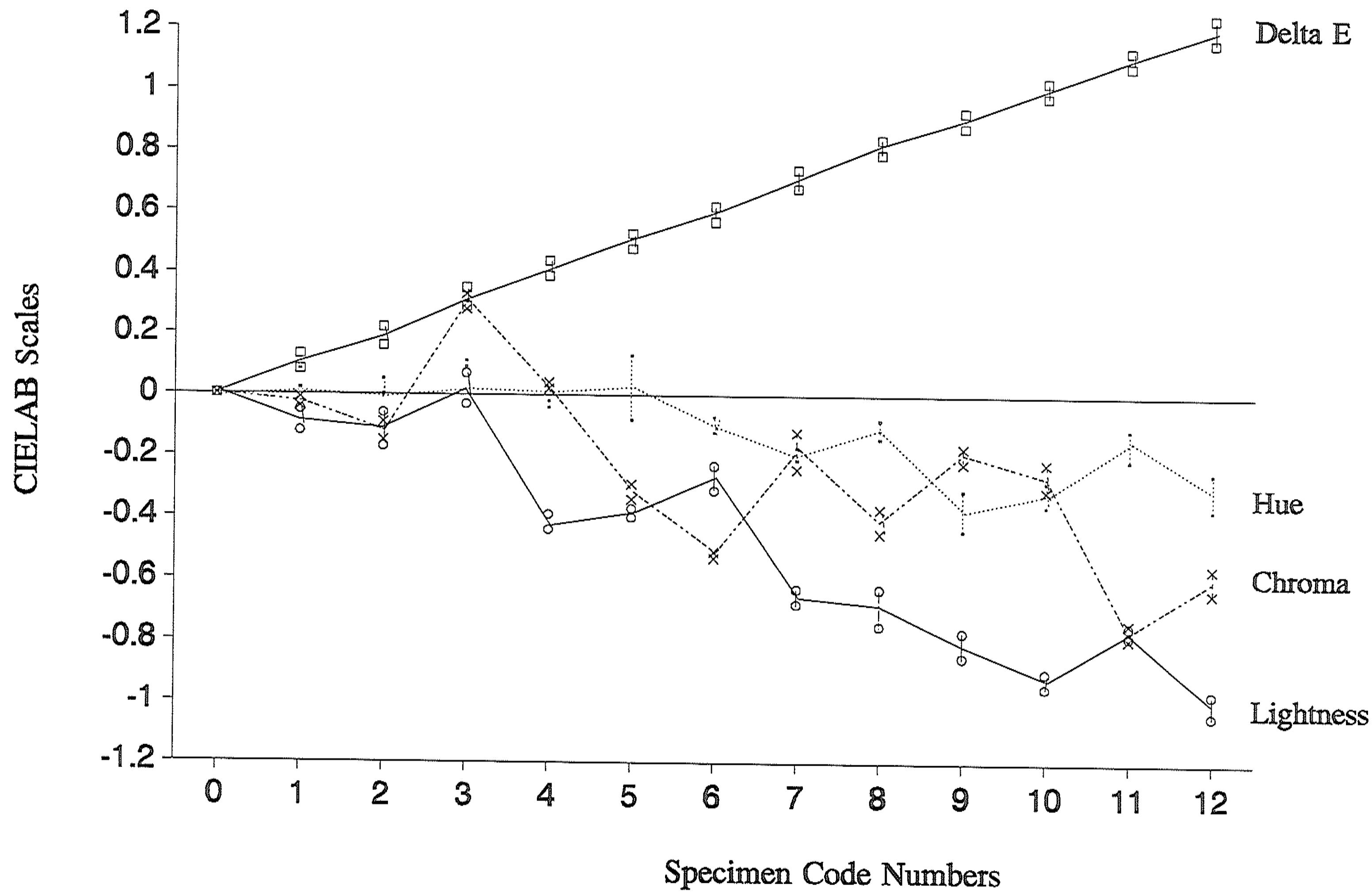

Figure 7. ΔE, ΔL, ΔC, and ΔH trend of selected specimens for the threshold test

hue were similar to the reference specimen, but it had higher chroma value (therefore brighter) than the reference specimen (Appendix H). The drastic change of chroma could have influenced subjects' responses in whiteness. These results supported the statement by Crown (1987) that the colour components lightness, chroma, and hue interact with each other in human visual perception. As the hue of specimen 3 was very similar to the reference specimen, it was not surprising that subjects could not detect any colour differences. Furthermore, because the question posed to the subjects was "which one appears yellowish white or bluish white", the subjects might not have seen yellow or blue hues.

The Ranking Test

The ranking test involved the selection of specimens from the pool of 102 dyed specimens. Forty two specimens met the criteria established earlier for ΔL, ΔC, and ΔH; 26 met the criteria of $\Delta E_{min} = 0.2$, $\Delta E_{med} = 0.5$, and $\Delta E_{max} = 0.8$ for visual ranking.

Specimen Selection

One criterion was that there had to be a reference specimen. The researcher selected the specimen with the instrument values of $L^* = 91.79$, $a^* = -0.39$, and $b^* = 2.27$. Then, ΔE, ΔL, ΔC, and ΔH for each of the 102 dyed specimens were calculated from this reference specimen. Averaged ΔE values were listed in ascending order from the reference specimen. Using the ΔE values established in threshold test, the researcher looked for specimens that showed increments of 0.2, 0.5, and 0.8 ΔE (Figure 8). Only the specimens in which the ΔE values had no overlap with values from consecutive specimens were selected. This procedure ensured that the whiteness level for each specimen was different from the others. Also, each specimen's whiteness was controlled within a small tolerance.

A two-way analysis of variance was conducted on the 42 specimens for ranking to determine the effects of specimen, orientation, and the interaction between specimen and orientation on ΔE. Each specimen had a significant contribution to the total variance of ΔE measurement ($F = 2009.89$, $p = 0.0001$). The orientation effect was not significant ($F = 2.44$, $p = 0.1204$), nor was the interaction between specimen and orientation ($F = 0.36$, $p = 0.9999$).

The researcher selected 26 specimens for the ranking tests (Appendix I). The ΔE, ΔL, ΔC, and ΔH of the 26 specimens were plotted in Figure 9. When using the specimen which had the least amount of dye as the reference (specimen 1), the ΔE became greater with the increments of the dyestuffs on each specimen. ΔL, ΔC, and ΔH were controlled so that ΔL was getting darker, ΔC was getting duller,and ΔH was getting bluer as discussed in visual threshold section; therefore, no single colour component (ΔL, ΔC or ΔH) exerted any irregular influence on the visual perception of ΔE. The researcher then determined the Yxy colour space values for the 26 specimens. The ranges of values were found as follows: $Y = 62.40$ to 80.30, $x = 0.3120$ to 0.3138, $y = 0.3176$ to 0.3209, and fell within the off-white colour as defined for this study.

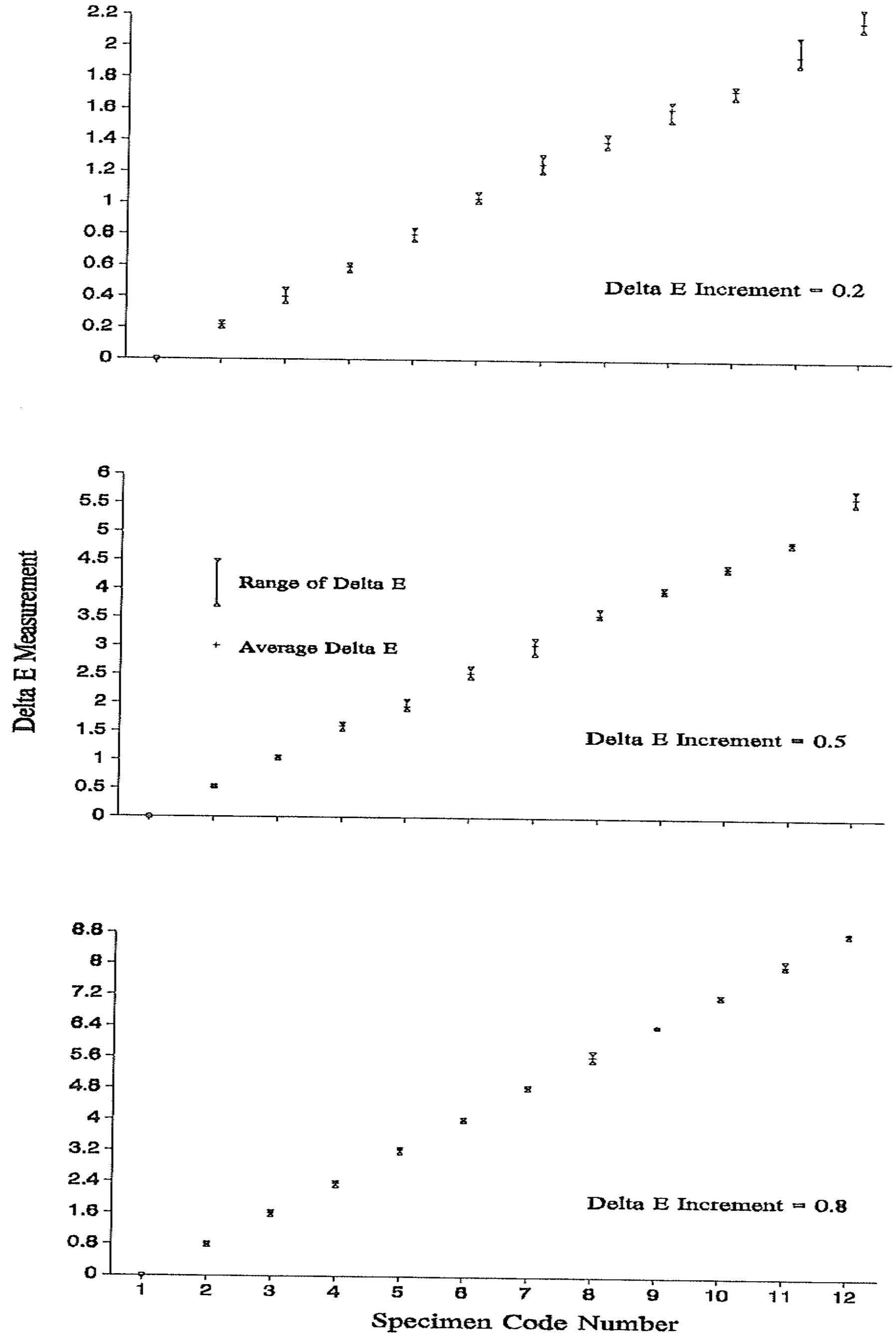

Figure 8. ΔE of selected specimens for ranking

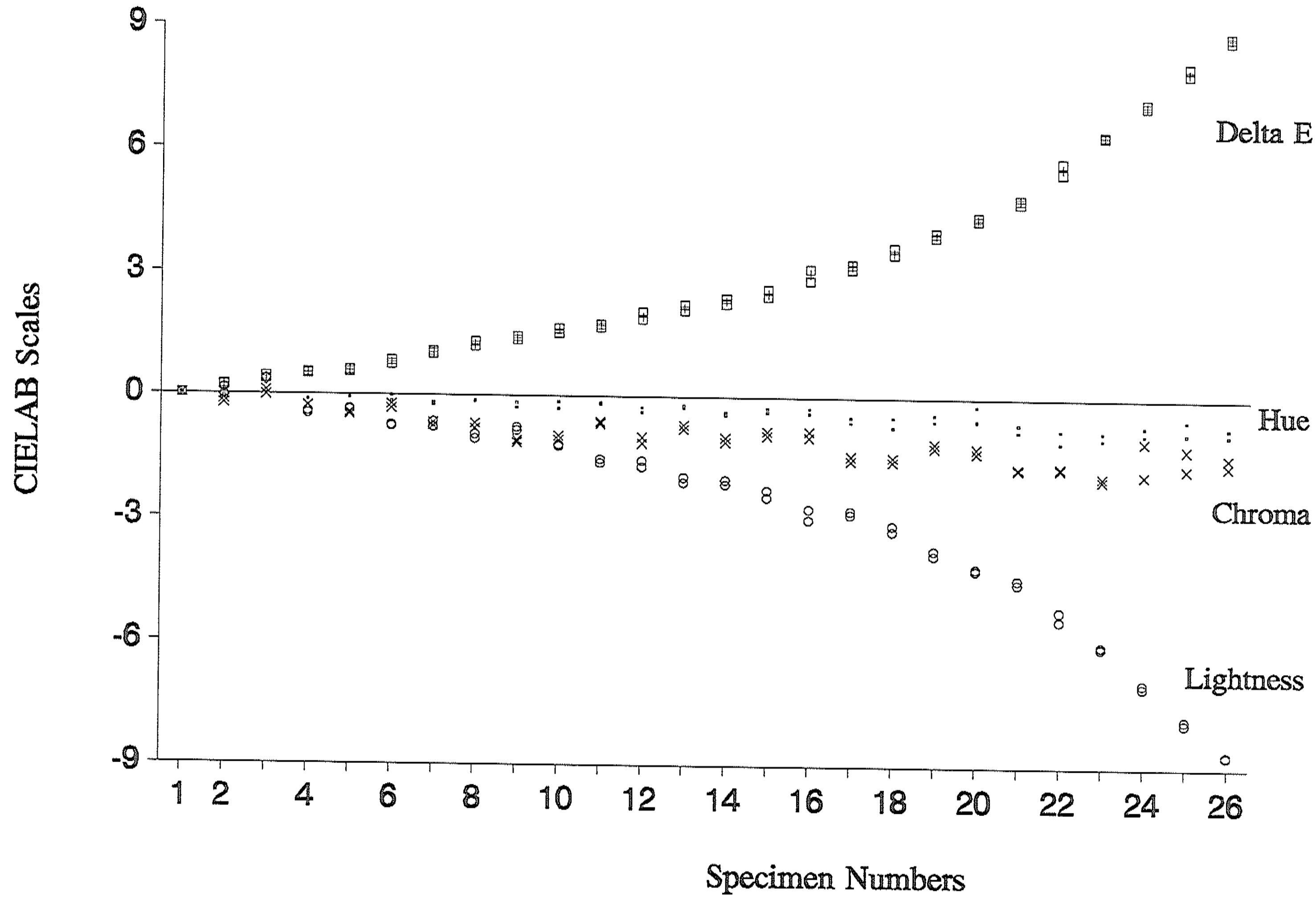

Figure 9. ΔE, ΔL, ΔC, and ΔH of selected specimens for ranking

Assessment of Visual Ranking Accuracy

Ranking was conducted under two factors using the randomized complete block design. After subjects ranked the specimens, the researcher had to calculate Kendall's Tau (τ) to test the first two null hypotheses. The τ value between the expected (instrument) and panelists' rank orders was used to determine the accuracy of ranking procedure for evaluating off-white fabrics. The $\hat{\tau}$ values for each subject are listed under the ΔE and number of specimens factors in Tables J-1 and J-2, respectively. Tables 6 and 7 list the median $\hat{\tau}$ values for the ΔE and number of specimen factors. The medians, instead of the averages, were used because this was a nonparametric analysis.

Table 6 shows that when the number of specimens were controlled median $\hat{\tau}$ values increased as ΔE increased. However, median $\hat{\tau}$ values did not decrease consistently as the number of specimens increased when the ΔE values were controlled (Table 7). When looking at the number of correct visual rankings ($\hat{\tau}$ = 1.00) in each column (Appendix J), the number of correct responses increased with increasing ΔE or stayed the same when the number of specimens were held constant. Also, the number of correct responses usually decreased with increasing number of specimens when ΔE was controlled. Errors in ranking different number of specimens could have influenced $\hat{\tau}$ values. For instance, a single reversal in three specimen series brought $\hat{\tau}$ value down to 0.33 while a single reversal in 12 specimen series produced $\hat{\tau}$ value of 0.97.

Table 6

Median $\hat{\tau}$ Values Collected From 24 Subjects Within Each Cell

(Ranking Done When ΔE Varied)

Number of specimens	ΔE		
	0.2	0.5	0.8
3	1.00	1.00	1.00
6	0.87	1.00	1.00
9	0.89	0.94	1.00
12	0.87	0.96	1.00

Table 7

Median $\hat{\tau}$ Values Collected From 24 Subjects Within Each Cell

(Ranking Done When Number of Specimens Varied)

ΔE	Number of specimens			
	3	6	9	12
0.2	1.00	0.73	0.83	0.77
0.5	1.00	1.00	0.94	0.96
0.8	1.00	1.00	1.00	0.99

Furthermore, the $\hat{\tau}$ values had to be ranked in each block under a given factor. The rank orders and rank sums are listed in Tables J-3 and J-4 in Appendix J, then Page's test is applied. The lowest value in each series was assigned one and the highest value was assigned three or four depending on the number of levels of a factor within each block. For the same $\hat{\tau}$ values in each block, the means of the rank positions (ties) were given. Table J-3 showed that the rank sums decreased with the increase of number of specimens when the ΔE values were controlled (i.e., fixed), and the rank sums increased with the increase of ΔE values when the number of specimens were fixed (Table J-4).

To answer the first research question, the null hypothesis that there is no difference in the accuracy of the visual ranking process when the ΔE of consecutive specimens changes from 0.2 to 0.8 was tested by Page's test (vs. Ha_1: The Kendall's Taus are ordered in the following way: $\tau_{0.2} \leq \tau_{0.5} \leq \tau_{0.8}$). Table 8 shows the results.

Table 8

P-Values for Testing Ho_1 vs Ha_1

(When ΔE Varied and Number of Specimens Was Constant)

	3	6	9	12
p	0.14	0.00	0.00	0.00

Except when ranking three specimens, all p-values are smaller than 0.05. Therefore, the null hypothesis was rejected when the number of specimens being ranked was 6, 9, and 12, meaning that the ranking accuracy was affected by the magnitude of ΔE between consecutive specimens where the number of specimens was controlled. The researcher was able to conclude that the experimental results were ordered as specified by the alternative hypothesis (Ha_1: The Kendall's Taus are ordered in the following way: $\tau_{0.2} \leq \tau_{0.5} \leq \tau_{0.8}$); therefore, the correlation between visual and instrument ranks increased as the ΔE between consecutive specimens increased.

To answer the second research question, the null hypothesis that there is no difference in the accuracy of the visual ranking process when the number in series of specimens being ranked varies from n = 3 to n = 12 was tested again by Page's test (vs. Ha_2: The Kendall's Taus are ordered in the following way: $\tau_3 \geq \tau_6 \geq \tau_9 \geq \tau_{12}$). Table 9 shows that all p-values are small.

Table 9

P-Values for Testing Ho_2 vs Ha_2

(When Number of Specimens Varied and ΔE Was Constant)

	0.2	0.5	0.8
p	0.000	0.005	0.026

Therefore, the null hypothesis was rejected. The researcher was able to conclude that the experimental results were ordered as specified by the alternative hypothesis (Ha_2: The Kendall's Taus are ordered in the following way: $\tau_3 \geq \tau_6 \geq \tau_9 \geq \tau_{12}$). This meant that the ranking accuracy was affected by the number of specimens being ranked at one time where the ΔE was controlled. The correlation between visual and instrument ranks, therefore, decreased as the number of specimens being ranked increased.

In addition to testing the null hypotheses, the researcher was interested in examining the relationship between the two factors. The relationship between the number of specimens and ΔE was investigated graphically. Figure 10 showed the relationship between these two factors. Under each factor (ΔE, and number of specimens), the median $\hat{\tau}$ values from subjects were plotted. The median absolute deviations were marked on each median (Figure 10). When evaluating three specimens, the panel could make accurate rankings ($\hat{\tau} = 1.00$) where the ΔEs between consecutive specimens were 0.2 ΔE units or higher. The panel produced excellent correlation coefficient when six specimens were ranked at one time where the ΔE between consecutive specimens was equal to or higher than 0.5 ΔE units.

Figure 10 illustrates even though the ranking data were collected under different factors and at different times over six weeks, the medians of $\hat{\tau}$ were in a similar range. The patterns when ΔE is 0.5 or 0.8 are similar but somewhat different from the pattern when ΔE equals 0.2. These results might be explained by the information from the threshold test. Every subject detected differences greater than 0.5 ΔE units during the threshold test, while one third had difficulties detecting differences of 0.2 ΔE units. Therefore, both of the detectable differences (0.5 and 0.8 ΔE units) generated high $\hat{\tau}$ values, and 0.8 ΔE differences produced slightly higher $\hat{\tau}$'s than did 0.5 ΔE. Figure 10 also showed that the higher the ΔE differences, the less variation of median $\hat{\tau}$'s was observed. There were some interactions between the two factors when less than nine specimens were ranked.

The third research question could not be addressed because there was insufficient data to calculate the threshold responses for each panelist. The 24 subjects were then divided into three groups according to the degree of

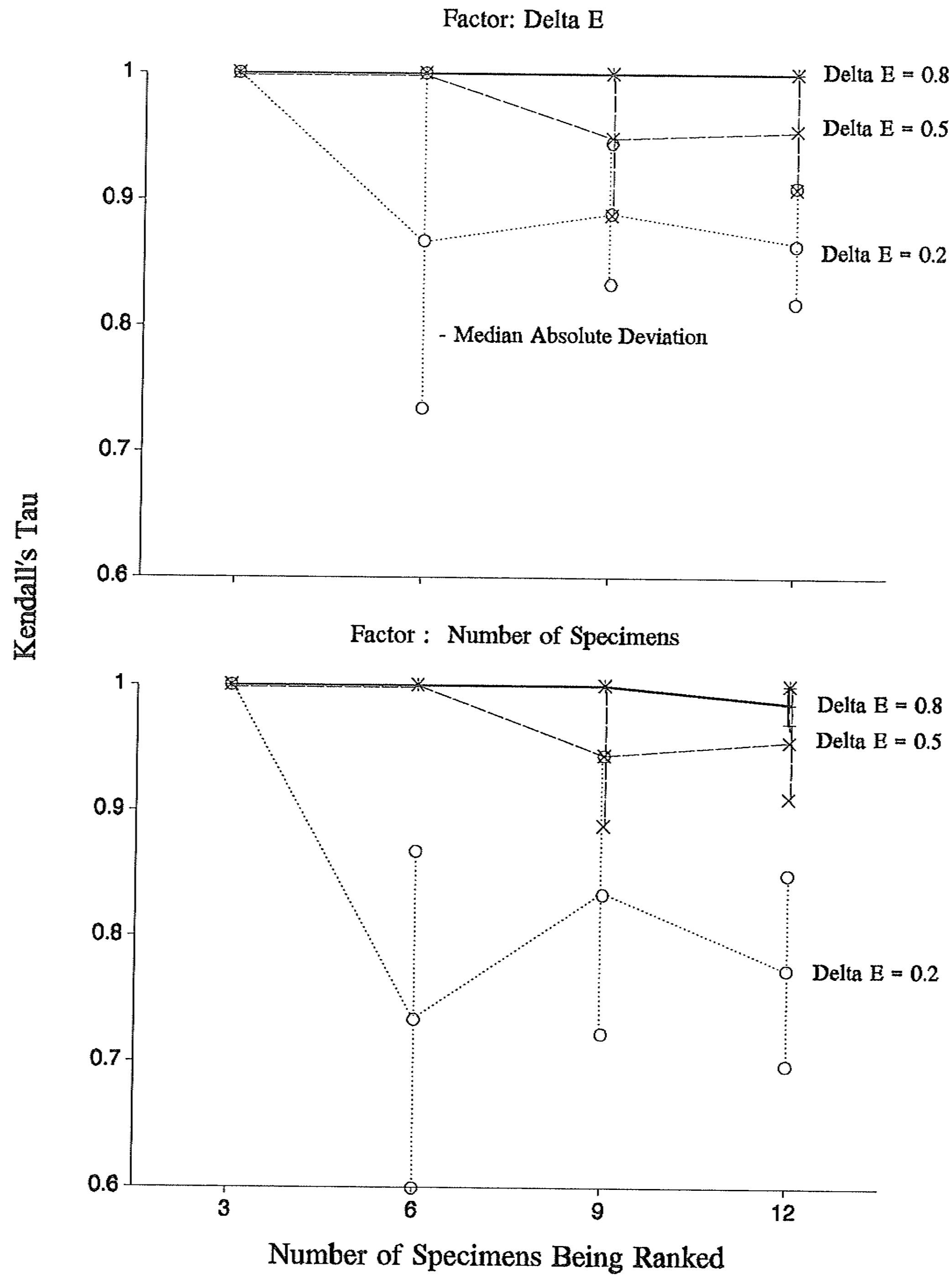

Figure 10. Relationship between median $\hat{\tau}$ and two factors

monotonicity for correct response curves. The groups were: (a) those whose correct response was always increasing with the increment of ΔE in the threshold test, (b) those with one drop from where the correct response should increase in the threshold test, (c) all else (Table H-2 in Appendix H). The percentages of correct responses were calculated for each subject from five threshold tests where ΔE ranged from 0.1 to 0.5 ΔE units. This range of ΔE was adopted because subjects had difficulty detecting colour differences in this range. The degree of monotonicity represented a level of visual perception. A relationship, therefore, between ranking accuracy and the level of visual perception can be examined by plotting median $\hat{\tau}$'s against percentage of correct responses grouped according to degree of monotonicity as shown in Figure 11. The median $\hat{\tau}$'s for this plot must be selected from the 0.1 to 0.5 ΔE range (i.e., 0.2). Median $\hat{\tau}$'s were chosen from the ranking factor where the number of specimens was equal to six because the subjects started to have difficulty in distinguishing the rank order from this point. The relationship still could not be fully explored because each group needed more subjects.

One expected that subjects with lower level of colour perception might also have lower ranking accuracy while higher level in perception might produce higher ranking accuracy. The plots for each group demonstrated different patterns which suggested that the level of visual perception could have an influence on ranking accuracy.

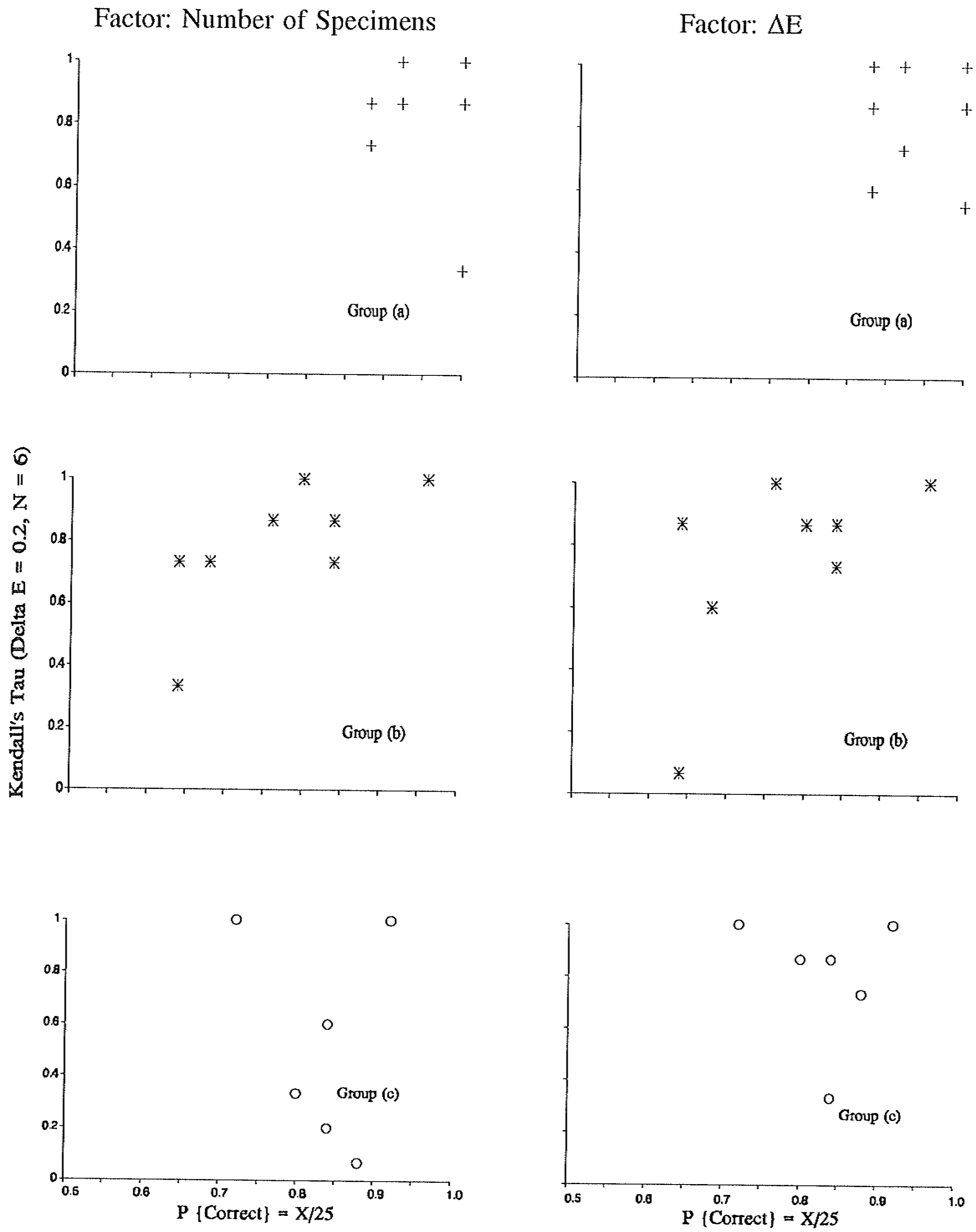

Figure 11. Relationship between $\hat{\tau}$ and possibility of correct responses in 25 paired comparisons by 24 subjects grouped according to degree of monotonicity of response curves under different factors.

CHAPTER 5

SUMMARY, CONCLUSION, AND IMPLICATIONS

Textile researchers have paid little attention to visual ranking and its relationship to instrument measurement in assessing fabric whiteness. This chapter summarizes an exploratory study to examine the accuracy of a visual ranking procedure. A series of experiments were conducted. Conclusions are drawn and suggestions for further research are outlined. Implications for visual sensory evaluation in product development are also discussed.

Summary

CIELAB ΔE is the only measurement specified in standard test methods to assess colour difference. However, studies investigating the relationship between CIELAB ΔE and visual ranking have not been located in textiles research. The researcher, therefore, had very little information to draw upon.

The researcher investigated the accuracy of a visual ranking method to assess colour differences in off-white fabrics under a series of controlled conditions. Ranking accuracy was estimated by the strength of the Kendall's Tau correlation between visual ranking and instrument ranking of the same series of specimens. Three research questions were proposed relating ranking accuracy to: (a) the magnitude of the colour difference between consecutive specimens; (b) the number of specimens being ranked; and (c) the level of subject's visual perception.

A series of preliminary experiments were conducted to verify the precision of the Hunterlab and to develop a computer program generating colorimetric values. The researcher produced a series of off-white specimens by first dyeing a standard white fabric with a given mixture of red, blue and yellow dyestuffs at different concentrations and dilutions. Then, the researcher selected specimens that had the same ΔE values between consecutive specimens. Specimen selection was based on the following criteria: (a) each specimen series was compared to one common reference specimen; (b) the ΔE measurements for each specimen were within a ± 0.04 range of the average; (c) the nominal ΔE measurement between consecutive specimens increased in equal increments from the whitest to the least white specimen; (d) ΔL, ΔC, and ΔH between consecutive specimens decreased from whitest to least white.

To derive the instrument values, the researcher developed a computer program to convert the CIE tristimulus readings X, Y, and Z into the 13 required colorimetric values. The computer program replaced the conventional method of collecting colorimetric instrument readings, produced values identical to each individual instrument measurement, minimized data collection time, facilitated data handling, reduced specimen contamination and minimized error resulting from the light source aging.

The specimens must be opaque because a direct comparison of visual and instrument rankings was required. The researcher adapted the mounting system reported in Vanderhoeven's study.

Twenty-four female students in Clothing and Textiles from the faculty of Human Ecology participated in the study in spring session of 1993. They were selected after being screened for colour deficiency.

The threshold test was conducted to establish the ΔE values for the colour difference factor in the ranking experiment. The ΔE values were determined as $\Delta E_{min} = 0.2$, $\Delta E_{med} = 0.5$, and $\Delta E_{max} = 0.8$. In the threshold test, the researcher found that interchanging the specimen positions improved the sensitivity of the visual threshold in the near white colour space.

Visual ranking data were collected using a randomized complete block design under two factors: (a) varying ΔE between consecutive specimens within a given series and (b) changing number of specimens being ranked at one time. The values (0.2, 0.5 and 0.8) for the ΔE factor were determined from the threshold test while the values and the increment (3, 6, 9 and 12) for the number of specimens factor were determined by the researcher.

The visual rankings produced ordinal measurements that required nonparametric statistical analyses. Kendall's Tau coefficient (τ) was calculated to determine the strength of relationship between each subject's ranking order and the expected ranking for each series within both factors. The expected rankings were established from the instrument measurements.

To address the first research question, Page's test for ordered alternatives was used to test the null hypothesis that there was no difference in the strength of relationship between visual and instrument ranking orders when the colour difference value between consecutive specimens increased from $\Delta E = 0.2$ to $\Delta E = 0.8$. The null hypothesis was rejected. The researcher was able to conclude that the experimental results were ordered as specified by the alternative hypothesis (Ha_1: the strength of relationship between visual and instrument rankings increases as the colour difference value of consecutive specimens increases).

To address the second research question, Page's test was also used to test the second null hypothesis that there was no difference in the strength of relationship between visual and instrument ranking orders when the number in series of specimens being ranked varied from $n = 3$ to $n = 12$. This null hypothesis was also rejected. The researcher was able to conclude that the experimental results were ordered as specified by the alternative hypothesis (Ha_2: the strength of relationship between visual and instrument rankings decreases as the number in series of specimens being ranked increases).

The third research question "Does a panelist's level of perception affect ranking accuracy?" was not addressed because there was not sufficient observations for statistical analysis; more comparisons were needed per subject in the threshold test. Therefore, level of colour perception could not be statistically related to ranking accuracy. Consequently, the researcher graphically examined the relationship between the level of colour perception and ranking accuracy. The researcher grouped correct responses from the threshold experiment according to their degrees of monotonicity and observed that the level of colour perception could have an influence on ranking accuracy.

Conclusions

Relationship Between Visual and Instrument Assessments

Although instruments generate quantitative data, they cannot totally replace human visual responses. Hunterlab used in this study could detect minute colour differences and generate interval data. Visual evaluation, however, can be used to directly test subjects' response and to help researchers interpret instrument data. The experimental design helped to reveal some numerical relations among accuracy of visual ranking, ΔE and number of specimens in this study.

The overall strength of relationship between visual and instrument rankings in this study was fairly high. The results illustrated that the median $\hat{\tau}$s for panel's visual ranking and standard instrument readings (ΔE) ranged from 0.73 to 1.00 when ΔE was equal to or greater than 0.2, and when the number of specimens being ranked varied from 3 to 12. It seems that the method of collecting the ranking data did not influence the median $\hat{\tau}$ values.

Some researchers indicated that ranking was an inaccurate method to assess whiteness while others supported it by reporting "good" correlations with instrument measurements. From this study, the researcher found that certain ranking conditions could produce acceptable sensory evaluations and might have potential use for visual evaluation of fabric whiteness.

Relationship Between Ranking Accuracy and the Two Factors

In this research, ranking accuracy increased as the ΔE between consecutive specimens increased and as the number of specimens being ranked decreased. When less than nine specimens were ranked, some interactions between the two factors were observed. Median $\hat{\tau}$ values of 1.00 were observed for all the rankings involving three specimens. Also, identical $\hat{\tau}$ values were observed when six or more specimens were ranked and where the ΔE between consecutive specimens was equal to or greater than 0.5 ΔE units. The median $\hat{\tau}$ only dropped from 1.00 to 0.73 when $\Delta E = 0.2$ and when six or more specimens were used; therefore, the researcher questioned the ranking accuracy where specimens with 0.2 ΔE units were involved.

This study supported the conclusion by Coppock (1965) that panelists were unable to rank the specimens when the colour difference was small and the number of specimens was excessive (30). The high correlation between instrument measurement and visual evaluation reported by Furry et al. (1961), Rhode Island Section (1966), and Warfield and Hardin (1981) might have resulted from easily detectable colour difference and small number of specimens. The researcher could not compare the influences on ranking accuracy from colour difference values reported from previous research mentioned above because the values were in different units.

The number of specimens being ranked as well as human's ability to detect minute colour difference influenced ranking accuracy. For example, a single reversal in a rank order for 3, 6, 9, or 12 specimens produced $\hat{\tau}$ values of 0.33, 0.87, 0.94, or 0.97, respectively.

Implications for Further Research

This research raised some further questions about ranking. The researcher, therefore, makes some suggestions for further research.

Improvement of Threshold Test

The threshold test needs to be conducted again. The distribution of correct responses in the threshold test of this study suggested that the visual threshold sensitivity could be improved by allowing the panelists to manipulate the specimens during the evaluation. To increase the precision and accuracy of observations, a higher number of replications for each ΔE and each subject is required. The researcher also suggests that larger number of screened subjects should be used in the panel if the result from threshold test will be used to compare with that from the ranking test again. The specimens' ΔE, ΔL, ΔC, and ΔH may be controlled according to the specimen criteria developed for the ranking experiments in this study. The results with and without interchanging specimens may be compared again.

The Point at Which Ranking Accuracy Declines

Results of this research showed that ranking accuracy was lowered when ΔE was 0.2 and number of specimens was greater than three. The factors ΔE and number of specimens, therefore, need to be refined to reveal the point at which the ranking accuracy begins to decline. In future research, the researcher suggests using specimens with ΔE ranging from 0.1 to 0.6 with an increment of 0.1 units and that the number of specimens ranges from three to nine with an increment of one specimen because the ranking accuracy was lowered in this area.

Ranking Accuracy When ΔE Is Not in Equal Increments

In this research, the ΔEs between consecutive specimens were in equal increments. In future study, researchers may examine the effects of ΔE and number of specimens on ranking accuracy when ΔEs are not in equal increments. Specifically, researchers may randomly select ranking series from a pool of specimens with ΔEs ranging from 0.1 to 0.6 and number of specimens ranging from three to nine. These unequal increments are closer to the practical situation in textiles.

Comparison of Ranking and Paired Comparison Methods

In textiles research, comparison of ranking and paired comparison methods has not been located. In this study, although both paired comparison and ranking tests were conducted, the researcher did not compare the two methods. In this research, the accuracy of the two methods could not be directly compared because the threshold test produced percentage of correct responses while the ranking test generated median τs. The researcher, therefore, could not conclude that one method was better or more sensitive than the other because each had different units of measurement. Accuracy of visual and instrument assessment needs to be investigated using paired comparison and ranking methods. The purpose is to compare the accuracy of the two methods under the same controlled conditions.

Benefits to Detergent Company

Two of the four Canadian detergent manufacturers contacted by the researcher in 1991 were using both visual and instrument methods in developing their products. This study may enhance detergent manufacturers' understanding of the relationship between visual and instrument methods. This study provides insights into designing and conducting paired comparison and ranking tests, and relating instrument measurements to visual evaluations.

The computer program derived for this study may be used by the companies when large numbers of specimen measurements are involved. The researcher would recommend this computer program for further studies involving colour measurement for acceptability because the computer program calculates DEcmc, a value which the current Hunterlab cannot generate.

Consumers of detergents expect a white textile product to retain its original colour. Paired comparison can be used to investigate consumers' acceptance of colour change because the method concentrates on colour differences between original and washed fabrics. Ranking can be used to investigate how well different detergent formulations retain whiteness because the method concentrates on colour differences between adjacent (washed) specimens. Detergent manufacturers may consider the ranking method to assess what level of colour differences or changes is perceivable by or acceptable to consumers.

References

AATCC Evaluation Procedure 1. (1990). Grey scale for colour change. AATCC Technical Manual, 65, 331-332.

AATCC Test Method 110. (1989). Whiteness of textiles. AATCC Technical Manual, 65, 167-168.

AATCC Test Method 153. (1985). Colour measurement of textiles: instrumental. AATCC Technical Manual, 65, 276-281.

AATCC Test Method 173. (1989). CMC: Calculation of small colour differences for acceptability. AATCC Technical Manual, 65, 327-330.

ASTM D 1729. (1989). Standard practice for visual evaluation of colour differences of opaque materials. Annual Book of ASTM Standards, 06.01, 246-248.

ASTM D 2244 (1989). Standard test method for calculation of colour differences from instrumentally measured colour coordinates. Annual Book of ASTM Standards, 06.01, 297-301.

ASTM E 284. (1991). Standard terminology of appearance. Annual Book of ASTM Standards, 14.02, 216-231.

ASTM E 313. (1987). Standard test method for indexes of whiteness and yellowness of near-white, opaque materials. Annual Book of ASTM Standards, 14.02, 216-219.

Billmeyer, F. W. Jr., & Saltzman, M. (1981). Principles of colour technology. New York: John Wiley & Sons.

Brown, D. M., Cameron, B. A., Meyer, S. S., & Umber, J. J. (1991). The effects of water hardness level on washing quality using commercial laundry detergents. Journal of Consumer Studies and Home Economics, 15, 215-222.

Buchanan, B., Givon, M., & Goldman, A. (1987). Measurement of discrimination ability in taste test: An empirical investigation. Journal of Marketing Research, 24, 154-163.

Burnham, R. W., Onley, J. W., & Witzel, R. F. (1970). Exploratory investigation of perceptual colour scaling. Journal of the Optical Society of America, 60, 1410-1420.

Carver, M. N., & Wylie, M. J. (1980). Fabric whiteness as affected by selected laundry treatments and environmental factors. Home Economics Research Journal, 9, 96-104.

Clemson Conference. (1990). Colour measurement: Visual and instrumental. Textile Chemist and Colourist, 22, 18-22.

Coppock, W. A. (1965). The chemstrand whiteness scale. American Dyestuff Reporter, 54, 343-348.

Crown, E. M. (1978). Consumer perception and acceptance of colour change in textiles. Home Economics Research Journal, 6, 206-211.

Cutler, W. G., & Davis, R. C. (1972). Detergency, theory and test methods. New York: Marcel Dekker, Inc.

Daniel, W. W. (1978). Applied nonparametric statistics. Boston: Houghton Mifflin Company.

Furry, M. S., Bensing, P. L., & Johnson, M. L. (1961). Colour effects produced in nearly white fabrics by household detergents. American Dyestuff Reporter, 50, 50-55.

Galbraith, R. L., Swartzlander, M., & Hardin, I. R. (1987). Consumer perception of changes in fabric colour. Canadian Apparel Manufacturer, 1, 37-47.

Harold, R. W. (1987). Textiles: Appearance analysis and shade sorting. Textile Chemist and Colourist, 19, 23-31.

Hoban, R. F. (1981). Problems associated with colour measurements, reflectance and transmittance. Textile Chemist and Colourist, 13, 112-114.

Hunter, R. S. (1961). Achieving accuracy in measurements of textiles for reflectance and whiteness. American Dyestuff Reporter, 50, 812-820.

Hunter, R. S. (1975). The measurement of appearance. New York: John Wiley & Sons.

Jakobi, G., & Lohr, A. (1987). Detergents and textile washing - principles and practices. New York: VCH Publishers.

Kim, J., Smith, B. F., & Spivak, S. M. (1987). Comparative study of phosphate and non-phosphate detergents. Journal of Consumer Studies and Home Economics, 11, 219-235.

Larmond, E. (1977). Laboratory methods for sensory evaluation of food. Ottawa, Canada: Canadian Government Publishing Centre.

Lovingood, R. P., Woodard, J. E., & Leech, I. E. (1989). Three types of laundry detergents: Performance in soft and hard water. Home Economics Research Journal, 17, 253-262.

Matthews, W. L. Jr. (1968). Preparations and mounting textile samples for colour measurement. American Dyestuff Reporter, 57, 584-590.

Meilgaard, M., Civille, G. V., & Carr, B. T. (1987). Sensory evaluation techniques. Florida: CRC Press.

Morris, M. A. (1970). Laundering cotton fabric, Part II: Effects of detergent type and water temperature on soil removal. Textile Research Journal, 40, 637-643.

Morris, M. A., & Prato, H. H. (1982). The effect of wash temperature on removal of particulate and oily soil from fabrics of varying fiber content. Textile Research Journal, 52, 280-286.

Morris. M. A., & Prato, H. H. (1985). Relationship of wash/rinse water temperatures to soil removal. Journal of Consumer Studies and Home Economics, 9, 73-79.

O'Mahony, M. (1986). Sensory evaluation of food (statistical methods and procedures). New York: Marcel Dekker, Inc.

Paek, S. L. (1983). The effect of detergent type and laundering temperature on the soil removal and soil redeposition performance of cotton and P/C blends with and without durable press finish. Journal of Consumer Studies and Home Economics, 7, 339-348.

Palmetto Section. (1988). The photometric accuracy and repeatability of several spectrophotometers. Textile Chemist and Colourist, 20, 17-22.

Pelton, W. R. (1989). Laundry detergent effectiveness - An independent study. Journal of the Home Economics Association of Australia, 21, 71-78.

Pelton, W. R. (1990). [Laundry detergent effectiveness]. Unpublished raw data.

Rhode Island Section. (1966). Instrumental evaluation of the colour of fluorescent textiles in term of average observer response. American Dyestuff Reporter, 55, 997-1002.

Robertson, A. R. (1986). Recent CIE work on colour difference evaluation. In J. J. Rennilson, & W. N. Hale (Ed.), Review and evaluation of appearance: methods and techniques, ASTM STP 914 (pp. 33-39). Philadelphia: ASTM.

Stearns, E. I. (1974). Measurement of colour and colour differences. Textile Chemist and Colourist, 6, 38-49.

Vanderhoeven, R. E. (1992). Conversion of a visual to an instrumental colour matching system: An exploratory approach. Textile Chemist and Colourist, 24, 19-25.

Warfield, C. L., & Hardin, I. R. (1981). Visual vs. instrumental evaluation of fabric whiteness. Textile Research Journal, 51, 725-731.

White, N. L., Prato, H. H., & Morris, M. A. (1984). Consumer attitudes towards level of soil removal and energy use in the laundering process. Journal of Consumer Studies and Home Economics, 8, 251-262.

Wilcock, A. E., & Van Delden, E. L. (1985). A study of the effects of repeated commercial laundering on the performance of 50/50 polyester/cotton momic cloth. Journal of Consumer Studies and Home Economics, 9, 275-281.

Wilson, P. A. (1987). The effect of repeated use of fabric softeners on absorbency, whiteness and hand of terry towels. Book of Papers, 291-296.

APPENDICES

Appendix A

Some Detailed Information from the Selected Research

Textile Applications of Instrument Measurement

Equipment

Photoelectric colour difference metre (Furry et al., 1961)
General electric recording spectrophotometer (Model G100) (Coppock, 1965)
Large sphere signature model colour eye (Rhode Island Section, 1966)
Hunter colour and colour difference meter (Carver & Wylie, 1980)
Hunterlab D-40 reflectometer (Warfield & Hardin, 1981)
Macbeth Spectrophotometer (MS -2000 Illuminant C) (Morris & Prato, 1982)
Hunterlab D-40 reflectometer (Paek, 1983)
Spectrophotometer with illuminant C (White et al., 1984)
Hunterlab D-25 colour difference meter (Wilcock & Van Delden, 1985)
Beckman spectrophotometer and Hunterlab colour difference meter (Galbraith et al., 1987)
Hunterlab colorimeter (Kim et al., 1987)
Hunterlab with D65 illuminant (Lovingood et al., 1989)

Specimen Discolouration

Soiled and laundered (Morris and Prato, 1982)
Soiled and laundered (Paek, 1983)
Soiled and laundered (Morris and Prato, 1985)
Soiled and laundered ((Brown et al., 1991)

Layers and/or Dimensions of Specimens

Two inches square (5.08 cm^2) (Furry et al., 1961)
Eight layers and 10 cm^2 (Warfield, & Hardin, 1981)
Four layers and four centimetre square (Morris & Prato, 1982)
Four layers (Paek, 1983)
Six by five inches (15.2 X 12.7 cm^2) (White et al., 1984)
Two layers (Morris & Prato, 1985)
15 X 12.5 cm^2 (Brown et al., 1991)

Handling of Specimens

Face side and warp direction

Textiles Applications of Visual Evaluation

Light Source

North daylight (UV + Macbeth Examolite - type C-4D-UV) (Furry et al., 1961)
Approximate daylight (Coppock, 1965)
Daylight (Macbeth Lamp model #BBX - 826 UV + Near UV) (Rhode Island Section, 1966)
North sky light (Morris, 1970)
Artificial daylight source (Kim et al., 1987)
Macbeth artificial daylight source (Wilson, 1987)
Xenon-arc light (Colour matching booth (Pelton, 1989)

Specimen Discolouration

Applying various blue, violet, green, and red pigments to shift purity and brightness values; adding carbon black to change brightness without seriously affecting purity (Coppock, 1965)
White cotton fabrics soiled and laundered (Morris, 1970)
White (polyester/cotton) fabric soiled and laundered; same difference between each adjacent specimens (Warfield and Hardin, 1981)
Soiled and laundered (White et al., 1984)
Discolourations of white fabrics by dyeing (Galbraith et al., 1987)
Soiled and laundered (Pelton, 1989)

Layers and/or Dimensions of specimens

Two layers and two inches square (5.08 cm^2) (Furry et al., 1961)
Four layers and four by six inches square (10.2 X 15.2 cm^2) (Rhode Island Section, 1966)
10 X 10 cm^2 (Warfield and Hardin, 1981)
Six by five inches (15.2 X 12.7 cm^2) (White et al., 1984)
Five inches square (12.7 cm^2) (Galbraith et al., 1987)

Number of Specimens

12 white unsoiled fabrics (Furry et al., 1961)
30 white dyed fabrics (Coppock, 1965)
10 bleached cotton fabrics with brighteners (Rhode Island Section, 1966)
10 white soiled and laundered fabrics (Warfield, & Hardin, 1981)
39 specimens in whiteness (Galbraith et al., 1987)
Eight laundered fabrics (Pelton, 1989)

Backing of Specimens

Light grey (Furry et al., 1961)

Neutral background (Brightness 60%) (Coppock, 1965)

A background of four layers of bleached cotton fabric (Rhode Island Section, 1966)

Plane grey surface (Morris, 1970)

Viewing board with a standard AATCC overhead lighting (Galbraith et al., 1987)

Macbeth colour chamber (Wilson, 1987)

Number of Panelists

Three untrained workers (Furry et al., 1961)

20 screened female panelists (Coppock, 1965)

54 different observers at five firms (Rhode Island Section, 1966)

Three panelists at each of the six laboratories (Morris, 1970)

Four trained panelists (Warfield, & Hardin, 1981)

40 panelists (White et al., 1984)

138 consumer panelists (Galbraith et al., 1987)

15 untrained panelists (Wilson, 1987)

15 untrained panelists (Pelton, 1989)

Screening of Panelists

An average correlation coefficient of more than 0.70 from four independent (paired comparison) rank series (Coppock, 1965).

Pseudo-isochromatic Plates for Testing Colour Perception (Galbraith et al., 1987)

Terminology Used for Assessment

Blueish white to creamy white (Furry et al., 1961)

Whitest to least white (Coppock, 1965)

Whitest to least white (Rhode Island Section, 1966)

Darkest to lightest (Warfield and Hardin, 1981)

Handling of Specimen

Manipulation of specimens was allowed (Rhode Island Section, 1966)

Orientation kept constant (Morris, 1970)

Appendix B

Instrument Reliability

The Hunterlab manual listed the L, a and b values for the standard white tile accompanying the instrument. The L, a and b values for the standard white tile were 90.94, -1.20 and 1.80, respectively (Hunterlab Manual).

Tests were conducted using the standard white tile to assess drift and precision. The tile's L, a, and b measurements were standardized (90.94, -1.20 and 1.80) and four additional readings were then recorded at 15 minute intervals without moving the tile. The entire procedure was repeated over the next two days. The data from these tests were included in Appendix B. The Hunterlab procedure specified a maximum drift of 0.1 units in the tile's L, a or b measurements over one hour. The maximum drift in L, a and b from the standardization point over these three tests was 0.02, 0.04 and 0.02, respectively (Table B-1). The maximum standard deviations for L, a and b were 0.009, 0.029 and 0.008, respectively. The results were well within the specified requirement of 0.1.

Instrument accuracy were determined from L, a and b readings for the entire set of standard tiles (i.e., white, pink, green, yellow, blue and grey). The accuracy was estimated from the root mean square deviation (0.7 units maximum) for the entire series of six tiles (Hunterlab Manual). The calculated root means square deviations for the set of standard tiles were 0.018 for L, 0.043 for a and 0.022 for b. These results from Table B-2 indicated that the accuracy was well within the specified criteria. These tests confirmed that the instrument performance was stable, and the measurements were within the tolerance levels for drift, precision and accuracy for the L, a, b colour space.

The instrument measurement, ΔE, required for the visual sensory evaluations was calculated from the CIELAB L*, a* and b* values. The drift in the ΔE measurements from the standard white was also recorded (Table B-1). The maximum ΔE drift for the standard white tile measured at four intervals over an one hour was 0.04.

The Hunterlab preliminary test indicated that the instrument had no problem in detecting a ΔE colour difference of 0.1 between a pair of specimens. The ΔE measurements in Appendix B-1 suggested that the instrument was more sensitive to colour difference than the human eye ($\Delta E = \pm 0.2$) (AATCC Evaluation Procedure 1, 1987; Hunter Manual).

Table B-1

Instrument Drift and Precision

Tile	Time	Reading				Standard deviation		
		L	a	b	ΔE	S_L	S_a	S_b
White	D1	90.94	-1.20	1.80	0.00	0.008	0.024	0.008
		90.94	-1.22	1.80	0.02			
		90.93	-1.22	1.81	0.02			
		90.93	-1.20	1.82	0.03			
		90.92	-1.16	1.81	0.04			
	D2	90.94	-1.20	1.80	0.00	0.008	0.029	0.007
		90.93	-1.22	1.81	0.02			
		90.92	-1.16	1.81	0.04			
		90.92	-1.20	1.81	0.02			
		90.93	-1.24	1.82	0.04			
	D3	90.94	-1.20	1.80	0.00	0.009	0.020	0.007
		90.93	-1.18	1.79	0.02			
		90.92	-1.16	1.80	0.04			
		90.92	-1.16	1.80	0.04			
		90.92	-1.20	1.81	0.02			

Note: D represents different days for measurements.

Table B-2
Instrument Accuracy

Tile	Time	Reading			Standard deviation		
		L	a	b	S_L	S_a	S_b
White	W1	90.94	-1.20	1.80	0.000	0.000	0.005
	W2	90.94	-1.20	1.80			
	W3	90.94	-1.20	1.79			
	W4	90.94	-1.20	1.80			
Pink	W1	71.34	22.76	9.04	0.022	0.050	0.031
	W2	71.39	22.85	9.10			
	W3	71.36	22.83	9.11			
	W4	71.38	22.75	9.08			
Green	W1	64.68	-16.06	7.10	0.024	0.048	0.017
	W2	64.73	-15.99	7.14			
	W3	64.68	-15.95	7.12			
	W4	64.71	-15.97	7.13			
Yellow	W1	77.50	-2.27	20.70	0.019	0.043	0.025
	W2	77.53	-2.19	20.74			
	W3	77.49	-2.20	20.74			
	W4	77.49	-2.17	20.76			
Blue	W1	65.63	-7.41	-10.61	0.018	0.048	0.013
	W2	65.64	-7.36	-10.62			
	W3	65.61	-7.31	-10.64			
	W4	65.60	-7.31	-10.63			
Grey	W1	32.92	-0.10	0.49	0.017	0.045	0.027
	W2	32.92	0.01	0.49			
	W3	32.95	-0.04	0.48			
	W4	32.91	-0.04	0.54			

Note: W represents different weeks for measurements.

Appendix C

Computer Program for Instrument Data

The conventional way of obtaining instrument readings from the Hunterlab (Model D25M - 9) has been to record the values, as required, one by one from the instrument. For this study, 13 instrument values L*, a*, b*, ΔE (CIE), ΔL (CIE), ΔC (CIE), ΔH (CIE), L, a, b, Y, x, and y were required for each measurement. Each specimen required six measurements. Making the measurements and recording the data for several hundred specimens would be very time consuming.

The researcher, therefore, devised a computer program (Table C-1) to elicit the 13 required instrument values from the three CIE tristimulus values X, Y, and Z. The technique used in Test Method 173 (AATCC, 1989) was adopted to confirm the accuracy of computer calculations. AATCC Test Method 173 listed the L, a and b values which were generated from the tristimulus values of X, Y and Z for a series of standard pairs (red, blue, yellow, green, and grey). The Clothing and Textiles Department did not have this particular set of standard tiles. The program's accuracy was checked by measuring the 13 required colour space measurements for the six standard tiles accompanying the Hunterlab and then these values were compared to those generated by the program from the X, Y and Z values of the same standard tiles. The differences ranging from +0.02 to -0.01 between these two sets of results, recorded in Table C-1 were consistent with the instrument drift (0.1) established earlier.

The advantages of this program which were briefly introduced in the method are outlined in greater detail here:

<u>Less time recording values</u> With the computer program, only 18 (3 values X 6 measures) instead of 78 (13 X 6) values were required for each specimen. Therefore, a fraction of the time was needed for both recording the data from the Hunterlab and entering the data into a computer file for analysis.

<u>Minimising specimen contamination</u> The computer program required fewer measurements on each specimen when determining colour differences between a reference specimen and each of the other specimens. Once the instrument reliability was confirmed during the specimen measurement, the specimens would not be measured over and over. The specimens, therefore, were handled much less during the instrument measurement phase and had less chance of contamination using the computer technique than those measured in the conventional way.

<u>Efficient data handling</u> All the data could be collected in one computer file. The researcher could select any specimen as the reference point. From the recorded X, Y, and Z readings, the computer program took less than one minute to calculate all the necessary instrument values. These data would otherwise take several hours (depending on the data size) to record and enter into a computer file under the

conventional method. (It should be noted that recording and typing might incur some mistakes.) The computer program listed data in a format that either a SAS or Lotus program could read directly for further analysis. The data collected could be manipulated easily with computer programs to quickly examine the data around a reference point to locate a series of specimens which would meet the given criteria for the visual sensory evaluations.

Aging of instrument light source The life span of the Hunterlab light source could vary between six and eight months. A new light source could produce more consistent readings than those generated from the older bulbs (Hunter Manual). The computer technique developed in this study could collect data much quicker than using the conventional approach. Therefore, the aging of a light bulb had less influence on the data.

The computer program is listed below:

```
 DIM FY(3),FX(3)
' CMC(L:C) COLOUR DIFFERENCE FORMULA
'####################################
'INPUT DATA AND PRINT RESULTS
'####################################
  INPUT "INPUT CMC(L:C) WEIGHTING FACTOR 'L', 'C' ", L, C
  OPEN "I", #1, "YU.IN"
  OPEN "O", #2, "YU.OUT"
  LINE INPUT #1, S$
  PRINT #2," CIEL CIEA CIEB DEcmc DeltE WI   DL   DC   DH   L   A
B  X0  Y0"
  INPUT #1, X(1), X(2), X(3), XN, YN, ZN
  GOSUB 30:L1=CL:A1=CA:B1=CB
  LINE INPUT #1, S$
10 IF EOF(1)<>0 GOTO 20
  INPUT #1, X(1), X(2), X(3)
  X0=X(1)/(X(1)+X(2)+X(3))
  Y0=X(2)/(X(1)+X(2)+X(3))
  WI=4.0*X(3)/1.18103-3*X(2)
  GOSUB 30:L2=CL:A2=CA:B2=CB
  GOSUB 40
  LINE INPUT #1,S$
  PRINT #2, USING "###.##";CL,CA,CB,DE,DE0,WI, USING
"####.##";DL0,USING "###.##";DC0,DH,L4,A,B,USING "###.####";X0,Y0
  GOTO 10
20 CLOSE 1
  CLOSE 2
  PRINT
```

```
  END
 ' ############################
 ' CALCULATE L,A,B VALUES
 ' ############################
30 X(1)=X(1)/XN:X(2)=X(2)/YN:X(3)=X(3)/ZN
  L4=10*SQR(X(2)*YN)
  A=17.5*(X(1)*XN/0.98041-X(2)*YN)/SQR(X(2)*YN)
  B=7.0*(X(2)*YN-X(3)*ZN/1.18103)/SQR(X(2)*YN)
  FOR I=1 TO 3
  IF X(I)<8.856E-03 THEN FX(I)=7.787*X(I) ELSE FX(I)=X(I)^(1/3)-16/116
  NEXT I
  CL=116*FX(2):CA=500*(FX(1)-FX(2)):CB=200*(FX(2)-FX(3))
  RETURN
'##########################################
'CALCULATE CMC COLOUR DIFFERENCE
'##########################################
40
DL=L2-L1:DL0=DL:C1=SQR(B1*B1+A1*A1):C2=SQR(B2*B2+A2*A2):DC=C2-C1
  S1=DL*DL+(A2-A1)*(A2-A1)+(B2-B1)*(B2-B1)
  DE0=SQR(S1)
  DH=0:AA=S1-DL*DL-DC*DC:IF AA <=0 THEN 50 ELSE DH=SQR(AA)
  DH0=DH*SGN(CA)*SGN(CB)
  DC0=DC
50 IF (A2*B2)=0 THEN 60 ELSE H2=180-SGN(B2)*90-ATN(A2/B2)*57.3
  GOTO 70
60 BB2=SGN(ABS(B2)):AA2=SGN(A2+B2)
  H2=90*(BB2-AA2+1)
70 IF (A1*B1)=0 THEN 80 ELSE H1=180-SGN(B1)*90-ATN(A1/B1)*57.3
  GOTO 90
80 BB1=SGN(ABS(B1)):AA1=SGN(A1+B1)
  H1=90*(BB1-AA1+1)
90 IF H1<=164 OR H1>=345 THEN 100 ELSE GOTO 110
100 T=.36+ABS(.4*COS((H1+35)/57.3)):GOTO 120
110 T=.56+ABS(.2*COS((H1+168)/57.3))
120 SL=.040975*L1/(1+.01765*L1):IF L1<16 THEN LET SL=.511
  SC=.0638*C1/(1+.0131*C1)+.638:F=SQR(C1^4/(C1^4+1900))
  SH=SC*(T*F+1-F)
  DL=DL/(L*SL):DC=DC/(C*SC):DH=DH/SH
  DA=H2-H1:IF DA<0 THEN DH=-DH
  DE=SQR(DL*DL+DC*DC+DH*DH)
  DH=DH*SH
130 RETURN
140 END
```

Table C-1

Differences Between Calculated and Recorded Data

Standard	Required values											
tiles	L*	a*	b*	ΔE	ΔL	ΔC	ΔH	L	a	b	x	y
White	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pink	0.00	+0.01	0.00	+0.02	0.00	+0.02	-0.01	0.00	+0.02	+0.01	0.00	0.00
Green	0.00	0.00	0.00	0.00	0.00	+0.01	-0.01	+0.01	+0.01	0.00	0.00	0.00
Yellow	0.00	-0.01	+0.01	+0.01	+0.01	+0.01	0.00	+0.01	-0.01	+0.02	0.00	0.00
Blue	0.00	-0.01	0.00	+0.01	0.00	+0.01	0.00	0.00	-0.01	0.00	0.00	0.00
Grey	0.00	+0.01	0.00	+0.01	-0.01	-0.01	-0.02	0.00	+0.01	-0.01	0.00	0.00

Note:
1. Difference = Calculated Value - Recorded Value.
2. The Maximum Difference: from -0.01 to +0.02.

Appendix D

Dyeing Condition for Producing Off-White Specimens

Reactive dye: Remazol Yellow GR Lot 39593
Remazol Brilliant Red BB Lot 50170
Remazol Brilliant Blue R Special 37034
(Manufacturer: Hoechst Canada Inc.
4045 Côte Vertu
Montréal, Québec H4R 1R6)

(Note: Dye solution had to be heated, stirred with magnetic stirrer and cooled to room temperature. Distilled water was added to bring the solution back to original volume because of vaporization.)

Liquor/fabric ratio: 100 : 1

Electrolite: Na_2SO_4 22.0 g/L
Na_2CO_3 10.0 g/L

(Note: In the literature, 50 g/L of $Na_2SO_4 \cdot 10H_2O$ and 10 g/L of Na_2CO_3 were recommended. However, only anhydrous was Na_2SO_4 available in the lab. The amount per liter is recalculated.)

Container: 10 large stainless steel containers were used, 100 stainless steel balls per container.

Procedures:

1. To produce off-white specimens, the appropriate amount of electrolite (5.5 g Na_2SO_4 and 2.5 g Na_2CO_3 in 250 ml dye solution) were placed in each of 10 containers with 100 stainless steel balls. Dye solution and electrolite in each container had to be mixed completely. All specimens had to be wetted out before being placed in each of the containers.
2. The containers were placed in the launder-ometer at 75 °C, running for 30 min. Specimens were then rinsed in distilled water at 25 °C, placed in Hydro-extractor and ironed dry.
3. The stainless steel containers and balls had to be cleaned thoroughly after use.

Appendix E

Data Showing the Influence of Number of Layers and Backing Colour on ΔE

Table E-1

Data Showing the Influence of Number of Layers and Backing Colour on ΔE

ΔE measurement		Number	ΔE measurement		Number
White tile	Black tile	of layers	White tile	Black tile	of layers
2.38	11.98	1	0.29	0.98	6
2.39	11.81	1	0.29	0.83	6
2.42	11.97	1	0.26	0.89	6
2.42	11.79	1	0.28	0.94	6
2.36	11.94	1	0.34	0.98	6
2.34	11.99	1	0.51	1.07	6
1.56	5.72	2	0.15	0.59	7
1.58	5.59	2	0.16	0.59	7
1.56	5.67	2	0.14	0.57	7
1.56	5.63	2	0.12	0.58	7
1.57	5.60	2	0.17	0.58	7
1.70	5.77	2	0.27	0.73	7
1.07	3.32	3	0.12	0.36	8
1.08	3.34	3	0.12	0.37	8
1.04	3.30	3	0.07	0.31	8
1.05	3.33	3	0.07	0.35	8
1.15	3.37	3	0.11	0.37	8
1.11	3.40	3	0.15	0.49	8
0.71	2.13	4	0.16	0.21	9
0.71	2.13	4	0.18	0.22	9
0.68	2.12	4	0.11	0.17	9
0.71	2.13	4	0.10	0.18	9
0.76	2.12	4	0.04	0.20	9
0.81	2.22	4	0.07	0.34	9
0.44	1.39	5	0.16	0.18	10
0.46	1.40	5	0.17	0.16	10
0.44	1.37	5	0.09	0.10	10
0.49	1.39	5	0.09	0.08	10
0.51	1.38	5	0.07	0.10	10
0.57	1.54	5	0.06	0.19	10

Appendix F

Documents Relating Subjects

Approval for Research Proposal Involving Human Subjects

Letter of Invitation

Consent Form

Questionnaire for Subject's Background

FACULTY OF HUMAN ECOLOGY

UNIVERSITY OF MANITOBA

APPROVAL FOR RESEARCH PROPOSAL INVOLVING HUMAN SUBJECTS

This is to certify that: Mr. Chenyu Yang, Department of Clothing and Textiles, of University of Manitoba

presented a proposal for a research project entitled:

A Study of Ranking Procedure for Evaluating Off-White Fabrics.

The Faculty of Human Ecology Ethics Review Committee is satisfied that the appropriate ethical criteria for research involving human subjects have been met.

Members of the Committee:

Name	Position	Department
D. Fitzpatrick	Associate Professor	Foods & Nutrition
C. Harvey	Professor	Family Studies
N. Fetterman	Associate Professor	Clothing and Textiles

Date: March 15, 1993 ____________ ________________

Rosemary Mills
Committee Chair

Letter of Invitation

THE UNIVERSITY OF MANITOBA

FACULTY OF HUMAN ECOLOGY

Department of Clothing and Textiles

Human Ecology Building
Winnipeg, Manitoba
Canada R3T 2N2
(204) 474-8065 Office
(204) 275-5299 FAX

March 7, 1993

Dear Student:

I am a graduate student in the Clothing and Textiles Department and I am planning to study some sensory evaluation procedures which can apply to visual assessments of textile fabrics. The Faculty's Ethics Committee has reviewed and approved the study. This letter will explain what your commitment would be and the tasks involved if you agree to participate.

Previous research has established that gender can influence colour perception in visual sensory evaluations. I am controlling the gender variable by inviting only female students to participate in the study.

In the experiment, you will be required for one session and may be asked to undertake a further four testing sessions. Each session will require approximately 20 minutes to complete the visual evaluations. Your task will include identifying a series of colour motives and rating fabric pairs for colour differences, and may involve you ranking 3 or more fabrics from the whitest to the least white specimen. By participating in the study, you will gain some experience in sensory evaluation and insight into how visual sensory measurements are made on textile fabrics.

I will collect some background information from you during the first testing session. The information that you provide will be used only to describe the panel's profile for the research and will be kept strictly confidential.

The testing will take place during March, in the Textiles Graduate Office, Room H501 of Duff Roblin Building.

Please confirm your participation in the study by completing 1) the consent form and 2) the schedule listing your preferred test times. The consent form should be returned to Chengyu Yang by March 15, 1993. Once confirmations are received you will be advised of your test schedule. If you have any questions, please call Chengyu, 474-9616 (on campus)

I hope that you can assist me in this study.

Sincerely yours,

Chengyu Yang,
Graduate Student

Dr. W. R. Pelton,
Thesis Advisor

Consent Form

I have read the letter and understood the tasks and commitments of participants in the proposed study. I understand that:

1. Only female students are being invited to participate in the study;
2. This study will take place over a one month period;
3. I will be asked some background information;
4. I am free to choose not to answer any questions;
5. I am free to withdraw from the study at any time;
6. All information generated from this study will be kept in strict confidence.

I agree to participate in the visual sensory investigation entitled " A study of ranking procedure for evaluating off-white fabrics" which assesses fabric whiteness under different sets of conditions.

Name (Please Print) ______________________

Address ______________________

____

____

Phone ______________________

Signature ______________________

Date ______________________

Please indicate in the chart below when you are free to participate the sensory testing <u>for the coming two weeks</u>. This information will assist me in co-ordinating the testing.

	Monday	Tuesday	Wednesday	Thursday	Friday
AM	______	______	______	______	______
PM	______	______	______	______	______
AM	______	______	______	______	______
PM	______	______	______	______	______

Questionnaire for Subject's Background

At the beginning of session one, you were given an identification number which you are to use to identify yourself in this study. If you have forgotten the number, please ask the investigator for your number and enter it here ______.

Please provide the following background information by filling in the details requested or circling the appropriate answer.

1. What courses have you taken in clothing and textiles ? (If none - go straight to the next question) (Course Name or Code Number)
2. Are you primarily responsible for doing your own laundry ? Yes No
3. How many predominately white apparel items have you purchased in the past six months? Many Some Few None
4. Do you pay close attention to colour when purchasing apparel? Yes No
5. Your age is: Under 20 21-25 26-30 31-35 36-40 Over 40

Appendix G

Documents for Visual Evaluation

Instructions for Threshold Test

Instructions for Ranking Test

Instructions for Threshold Test

According to your perception, evaluate the whiteness of each pair of mounted fabric specimens, and answer the following questions:

A. "Can you see a colour difference between these specimens?" If the answer is "no", tell the researcher "there is no difference".

The researcher will record the response on a pre-printed table under "NO".

B. If the answer is "YES", indicate "which one of the pair appears yellowish white?" and "which one appears bluish white?" by pointing out the yellowish white and then to the bluish white specimen to the researcher.

The researcher will record the responses on the pre-printed table under "YES".

You will be allowed to alternate the position of each specimen pair during the exercise.

The researcher will show you an example of yellowish white and bluish white under the viewing condition before you commence.

During the evaluation, the researcher will not be able to answer any questions. If you have any questions, please ask them now.

If you do not have any questions, please start.

Instructions for Ranking Test

1. Between the two reference specimens placed in the Macbeth booth, you will rank the white fabrics in the order of whiteness according to their intensity from left to right (e.g., from the whitest to the least white).

2. When handling specimens, you will wear plastic gloves.

3. If two specimens appear the same, make a "best guess" of their rank order.

4. Once you have established the order from whitest to the least white, the researcher will record the rank order.

5. During the test, the researcher will not answer any questions. If you have any questions, please ask now.

6. If you do not have any questions, we will start.

Appendix H

Data Collected from Threshold Test

Table H-1

Colorimetric Data Collected from the Threshold Test

ΔE	ΔL	ΔC	ΔH	ID
1.18	-0.97	-0.62	-0.25	12
1.19	-0.96	-0.64	-0.28	12
1.21	-0.98	-0.60	-0.37	12
1.16	-0.98	-0.56	-0.28	12
1.24	-1.03	-0.60	-0.32	12
1.23	-1.04	-0.62	-0.22	12
1.09	-0.78	-0.74	-0.21	11
1.13	-0.78	-0.79	-0.21	11
1.10	-0.76	-0.78	-0.14	11
1.08	-0.75	-0.76	-0.17	11
1.10	-0.76	-0.79	-0.11	11
1.10	-0.77	-0.78	-0.14	11
1.03	-0.94	-0.25	-0.34	10
1.03	-0.93	-0.26	-0.36	10
1.00	-0.93	-0.24	-0.28	10
1.01	-0.95	-0.22	-0.28	10
1.01	-0.91	-0.31	-0.30	10
0.98	-0.90	-0.28	-0.28	10
0.89	-0.77	-0.19	-0.40	9
0.92	-0.78	-0.21	-0.44	9
0.90	-0.83	-0.18	-0.31	9
0.88	-0.80	-0.17	-0.31	9
0.91	-0.81	-0.21	-0.35	9
0.93	-0.85	-0.22	-0.31	9
0.84	-0.75	-0.37	-0.11	8
0.81	-0.71	-0.38	-0.09	8
0.79	-0.69	-0.39	-0.09	8
0.86	-0.74	-0.42	-0.08	8
0.79	-0.63	-0.45	-0.14	8
0.79	-0.65	-0.41	-0.14	8
0.74	-0.68	-0.22	-0.19	7
0.73	-0.67	-0.23	-0.17	7
0.70	-0.63	-0.24	-0.18	7
0.72	-0.65	-0.24	-0.20	7
0.68	-0.64	-0.12	-0.21	7
0.69	-0.65	-0.14	-0.18	7
0.60	-0.31	-0.51	-0.07	6
0.60	-0.29	-0.51	-0.12	6
0.58	-0.26	-0.51	-0.12	6
0.58	-0.26	-0.51	-0.12	6
0.57	-0.23	-0.52	-0.10	6
0.62	-0.30	-0.53	-0.12	6
0.48	-0.37	-0.31	0.04	5
0.49	-0.37	-0.30	0.11	5
0.48	-0.38	-0.30	0.04	5
0.49	-0.37	-0.29	0.13	5
0.52	-0.39	-0.34	-0.06	5
0.53	-0.40	-0.34	-0.08	5
0.39	-0.39	0.03	-0.02	4
0.43	-0.42	0.04	-0.02	4
0.40	-0.39	0.03	-0.04	4
0.43	-0.43	0.02	-0.04	4
0.44	-0.44	0.04	-0.02	4
0.43	-0.42	0.03	-0.02	4
0.30	-0.03	0.28	0.11	3
0.29	-0.01	0.28	0.09	3
0.34	0.05	0.32	0.11	3
0.31	0.00	0.29	0.10	3
0.35	0.06	0.33	0.09	3
0.35	0.07	0.33	0.09	3
0.22	-0.17	-0.13	0.02	2
0.20	-0.15	-0.12	0.04	2
0.16	-0.06	-0.15	0.01	2
0.18	-0.09	-0.15	0.04	2
0.19	-0.13	-0.13	-0.00	2
0.19	-0.16	-0.09	0.05	2
0.10	-0.06	-0.01	0.07	1
0.10	-0.07	-0.02	0.08	1
0.11	-0.08	-0.01	0.07	1
0.08	-0.05	-0.01	0.05	1
0.11	-0.10	-0.04	0.02	1
0.13	-0.12	-0.04	0.02	1
0.02	-0.02	0.00	0.01	0
0.03	0.02	-0.00	0.03	0
0.04	0.04	-0.01	-0.02	0
0.06	0.05	-0.02	-0.02	0
0.03	-0.03	0.02	0.01	0
0.06	-0.06	0.01	-0.01	0

Table H-2
Correct Responses from the Threshold Test

ΔE	Subject 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1.2	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
1.1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
1.0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
0.9	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
0.8	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
0.7	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
0.6	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
0.5	*	*	*	*	*	*	*	*	*	*	*	*	*	*	4	*	*	*	*	*	*	*	*	*
0.4	4	*	*	*	*	*	*	3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	4
0.3	*	*	3	2	1	3	*	*	*	2	3	4	3	*	3	3	*	*	3	*	4	*	*	*
0.2	*	*	*	3	*	*	*	*	*	3	*	*	4	4	3	4	*	*	*	*	*	*	4	3
0.1	1	2	3	2	4	0	3	*	*	1	4	2	4	3	1	4	2	*	3	3	*	*	3	2
Group	b	a	c	b	c	c	a	c	a	b	c	b	b	a	b	b	a	a	c	a	b	a	a	b

Note: "*" indicates 100% correct responses from five replicated tests.
Numbers indicate the number of correct responses from the five tests.

Appendix I

ΔE Values Selected for Ranking Series

Table I-1
ΔE Values Selected for Ranking Series

Order	ΔE = 0.2 Average	ΔE = 0.2 Range	ΔE = 0.5 Average	ΔE = 0.5 Range	ΔE = 0.8 Average	ΔE = 0.8 Range
1	**0.04**	0.02-0.06	0.04	0.02-0.06	0.04	0.02-0.06
2	**0.22**	0.20-0.24	**0.53**	0.51-0.54	0.80	0.76-0.84
3	**0.40**	0.36-0.45	1.03	1.01-1.07	1.61	1.53-1.65
4	**0.59**	0.56-0.61	1.61	1.53-1.65	**2.37**	2.29-2.42
5	**0.80**	0.76-0.84	1.95	1.89-2.07	**3.23**	3.15-3.27
6	**1.03**	1.01-1.07	**2.54**	2.46-2.64	3.99	3.96-4.04
7	**1.25**	1.20-1.31	**3.04**	2.87-3.14	4.85	4.79-4.87
8	**1.40**	1.36-1.44	**3.56**	3.52-3.67	5.64	5.51-5.77
9	**1.61**	1.53-1.65	**3.99**	3.96-4.04	**6.41**	6.39-6.43
10	**1.73**	1.68-1.75	**4.39**	4.35-4.44	**7.17**	7.14-7.22
11	**1.95**	1.89-2.07	**4.85**	4.79-4.87	**7.99**	7.94-8.11
12	**2.17**	2.12-2.25	**5.64**	5.51-5.77	**8.82**	8.76-8.84

Note: Bold numbers hilighted 26 specimens for all rankings.

Appendix J

Kendall's Tau Values from the Ranking Test under Different Factors

Table J-1

Kendall's Tau Correlation Coefficient from the Ranking Test (Factor: ΔE)

Sub.	ΔE = 0.2				ΔE = 0.5				ΔE = 0.8			
	Number of specimens				Number of specimens				Number of specimens			
	3	6	9	12	3	6	9	12	3	6	9	12
1	1.000	0.867	0.833	0.788	1.000	0.867	0.833	0.818	1.000	1.000	1.000	0.909
2	1.000	1.000	0.778	0.758	1.000	1.000	0.889	0.818	1.000	1.000	1.000	0.970
3	0.333	0.333	0.388	0.849	1.000	0.867	0.889	0.939	1.000	1.000	1.000	0.970
4	1.000	0.600	0.889	0.909	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1.000	0.867	0.778	0.818	1.000	1.000	1.000	0.970	1.000	1.000	1.000	0.970
6	1.000	1.000	0.944	0.909	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1.000	0.733	0.833	0.788	1.000	1.000	1.000	0.909	1.000	1.000	1.000	1.000
8	1.000	1.000	1.000	0.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	1.000	1.000	1.000	0.879	1.000	1.000	0.944	0.970	1.000	1.000	0.944	1.000
10	1.000	0.867	0.889	0.879	0.333	1.000	0.778	0.970	1.000	1.000	1.000	0.939
11	0.333	0.733	0.333	0.667	1.000	1.000	0.722	0.909	1.000	1.000	1.000	1.000
12	1.000	0.733	0.889	0.818	0.333	0.867	0.944	1.000	1.000	1.000	1.000	1.000
13	1.000	0.867	0.889	0.879	1.000	0.867	0.778	0.909	1.000	1.000	1.000	1.000
14	1.000	0.600	0.944	0.879	1.000	1.000	1.000	0.970	1.000	1.000	1.000	1.000
15	1.000	0.067	0.611	0.515	0.333	0.867	0.611	0.939	1.000	1.000	1.000	1.000
16	1.000	0.867	0.889	0.879	1.000	0.867	0.889	0.970	1.000	1.000	1.000	1.000
17	1.000	0.867	0.389	0.303	1.000	0.867	0.944	0.939	1.000	1.000	0.944	0.970
18	1.000	0.552	0.278	0.697	1.000	0.333	0.944	0.909	1.000	1.000	1.000	0.970
19	1.000	0.867	0.667	0.546	1.000	1.000	1.000	0.788	1.000	0.867	1.000	1.000
20	1.000	1.000	0.889	0.697	1.000	1.000	0.889	0.909	1.000	1.000	1.000	0.909
21	1.000	1.000	0.889	0.909	1.000	1.000	1.000	0.970	1.000	1.000	1.000	1.000
22	1.000	0.867	0.944	0.939	0.333	1.000	1.000	0.970	1.000	1.000	1.000	1.000
23	1.000	0.867	0.833	0.909	1.000	0.867	0.944	0.970	1.000	1.000	1.000	0.970
24	1.000	1.000	0.889	0.909	1.000	1.000	0.944	0.909	1.000	1.000	1.000	1.000
$\bar{\tau}_{Median}$	1.000	0.867	0.889	0.869	1.000	1.000	0.944	0.955	1.000	1.000	1.000	1.000

Table J-2
Kendall's Tau Correlation Coefficient from the Ranking Test (Factor: Number of Specimens)

Sub.	3			6			9			12		
	ΔE values			ΔE values			ΔE values			ΔE values		
	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
1	1.000	1.000	1.000	1.000	0.733	1.000	0.778	0.778	1.000	0.758	0.849	0.940
2	1.000	1.000	1.000	0.867	0.733	1.000	0.833	0.889	1.000	0.818	0.909	0.939
3	0.333	1.000	1.000	0.200	0.867	1.000	0.611	0.889	1.000	0.788	0.879	0.970
4	1.000	1.000	1.000	0.733	1.000	1.000	0.833	1.000	1.000	0.849	1.000	1.000
5	-0.333	1.000	1.000	0.333	1.000	1.000	0.667	0.944	0.944	0.818	1.000	0.970
6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.818	0.970	1.000
7	1.000	1.000	1.000	0.733	1.000	1.000	0.444	1.000	1.000	0.758	0.939	1.000
8	1.000	1.000	1.000	1.000	0.867	1.000	1.000	1.000	1.000	0.909	1.000	1.000
9	1.000	1.000	1.000	0.867	0.867	1.000	1.000	1.000	1.000	0.727	1.000	1.000
10	1.000	1.000	1.000	0.733	1.000	1.000	0.944	0.833	1.000	0.697	0.970	0.970
11	-0.333	1.000	1.000	0.067	1.000	1.000	-0.056	0.667	1.000	0.424	0.939	0.939
12	1.000	1.000	1.000	0.733	0.867	1.000	0.778	1.000	1.000	0.485	1.000	1.000
13	0.333	0.333	1.000	0.733	0.733	1.000	0.833	0.889	1.000	0.515	0.909	1.000
14	0.333	1.000	1.000	0.733	1.000	0.867	0.944	1.000	1.000	0.818	0.970	1.000
15	1.000	0.333	1.000	0.333	1.000	1.000	0.423	0.778	1.000	0.212	0.909	1.000
16	1.000	1.000	1.000	0.867	0.867	1.000	0.944	0.944	1.000	0.849	1.000	1.000
17	1.000	0.333	1.000	0.733	0.867	1.000	-0.278	1.000	0.944	0.394	0.939	0.939
18	1.000	1.000	1.000	0.333	0.333	1.000	-0.056	0.944	1.000	0.606	0.970	1.000
19	1.000	1.000	1.000	0.600	1.000	1.000	0.833	0.944	1.000	0.727	0.909	0.970
20	1.000	1.000	1.000	0.867	1.000	1.000	0.667	0.889	1.000	0.727	0.939	0.909
21	1.000	1.000	1.000	1.000	1.000	1.000	0.889	1.000	1.000	0.818	1.000	1.000
22	1.000	1.000	1.000	1.000	1.000	1.000	0.833	0.944	1.000	0.879	0.970	0.970
23	1.000	1.000	1.000	0.867	0.733	1.000	0.889	0.944	0.944	0.909	0.939	0.970
24	1.000	1.000	1.000	0.867	1.000	1.000	0.889	1.000	1.000	0.909	0.849	0.970
$\hat{\tau}_{Median}$	1.000	1.000	1.000	0.733	1.000	1.000	0.833	0.944	1.000	0.773	0.955	0.985

Table J-3
Ranks of Kendall's Tau Correlation Coefficient (Factor: ΔE)

Sub.	ΔE = 0.2				ΔE = 0.5				ΔE = 0.8			
	Number of specimens				Number of specimens				Number of specimens			
	3	6	9	12	3	6	9	12	3	6	9	12
1	4	3	2	1	4	3	2	1	3	3	3	1
2	3.5	3.5	2	1	3.5	3.5	2	1	3	3	3	1
3	1.5	1.5	3	4	4	1	2	3	3	3	3	1
4	4	1	2	3	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
5	4	3	1	2	3	3	3	1	3	3	3	1
6	3.5	3.5	2	1	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
7	4	1	2	3	3	3	3	1	2.5	2.5	2.5	2.5
8	3	3	3	1	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
9	3	3	3	1	3.5	3.5	1	2	3	3	1	3
10	4	1	3	2	1	4	2	3	3	3	3	1
11	3.5	2	3.5	1	3.5	3.5	1	2	2.5	2.5	2.5	2.5
12	4	1	3	2	1	2	3	4	2.5	2.5	2.5	2.5
13	4	1	3	2	4	2	1	3	2.5	2.5	2.5	2.5
14	4	1	3	2	3	3	3	1	2.5	2.5	2.5	2.5
15	4	1	3	2	1	3	2	4	2.5	2.5	2.5	2.5
16	4	1	3	2	4	1	2	3	2.5	2.5	2.5	2.5
17	4	3	2	1	4	1	3	2	3.5	3.5	1	2
18	4	2	1	3	4	1	3	2	3	3	3	1
19	4	3	2	1	3	3	3	1	3	1	3	3
20	3.5	3.5	2	1	3.5	3.5	1	2	3	3	3	1
21	3.5	3.5	1	2	3	3	3	1	2.5	2.5	2.5	2.5
22	4	1	3	2	1	3.5	3.5	2	2.5	2.5	2.5	2.5
23	4	2	1	3	4	1	2	3	3	3	3	1
24	3.5	3.5	1	2	3.5	3.5	2	1	2.5	2.5	2.5	2.5
Σ	88.5	52	54.5	45	72	62.5	55	50.5	66	64	61.5	48.5

Table J-4
Ranks of Kendall's Tau Correlation Coefficient (Factor: Number of Specimens)

S	3			6			9			12		
u	ΔE values			ΔE values			ΔE values			ΔE values		
b.	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
1	2	2	2	2.5	1	2.5	1.5	1.5	3	1	2	3
2	2	2	2	2	1	3	1	2	3	1	2	3
3	1	2.5	2.5	1	2	3	1	2	3	1	2	3
4	2	2	2	1	2.5	2.5	1	2.5	2.5	1	2.5	2.5
5	1	2.5	2.5	1	2.5	2.5	1	2.5	2.5	1	3	2
6	2	2	2	2	2	2	2	2	2	1	2	3
7	2	2	2	1	2.5	2.5	1	2.5	2.5	1	2	3
8	2	2	2	2.5	1	2.5	2	2	2	1	2.5	2.5
9	2	2	2	1.5	1.5	3	2	2	2	1	2.5	2.5
10	2	2	2	1	2.5	2.5	2	1	3	1	2.5	2.5
11	1	2.5	2.5	1	2.5	2.5	1	2	3	1	2.5	2.5
12	2	2	2	1	2	3	1	2.5	2.5	1	2.5	2.5
13	1.5	1.5	3	1.5	1.5	3	1	2	3	1	2	3
14	1	2.5	2.5	1	3	2	1	2.5	2.5	1	2	3
15	2.5	1	2.5	1	2.5	2.5	1	2	3	1	2	3
16	2	2	2	1.5	1.5	3	1.5	1.5	3	1	1.5	1.5
17	2.5	1	2.5	1	2	3	1	3	2	1	2.5	2.5
18	2	2	2	1.5	1.5	3	1	2	3	1	2	3
19	2	2	2	1	2.5	2.5	1	2	3	1	2	3
20	2	2	2	1	2.5	2.5	1	2	3	1	3	2
21	2	2	2	2	2	2	1	2.5	2.5	1	2.5	2.5
22	2	2	2	2	2	2	1	2	3	1	2.5	2.5
23	2	2	2	2	1	3	1	2.5	2.5	1	2	3
24	2	2	2	1	2.5	2.5	1	2.5	2.5	2	1	3
Σ	44.5	47.5	52	34	47.5	62.5	29	51	64	25	53	64