DOCTOR OF PHILOSOPHY

Children's perceptions of effort during cycling exercise

Lamb, K.L.

Award date:
1997

Awarding institution:
Bangor University

U

CHILDREN'S PERCEPTIONS OF EFFORT DURING CYCLING EXERCISE

K.L. Lamb B.A., B.Sc., M.Sc.

Thesis submitted for the Degree of Doctor of Philosophy of the University of Wales

School of Sport, Health and Physical Education Sciences,
University of Wales, Bangor

Spring 1997

CONTENTS

FIGURES

TABLES

ACKNOWLEDGEMENTS

I would like to recognize the invaluable contribution of the many participating school-children and their teachers to the development of this thesis. Thanks to Dr. J.G. Williams for his innovative work and consequent inspiration, and to my supervisor and friend, Dr. Roger Eston for his guidance and persistent confidence in my academic ability. Special thanks to my parents and family for their eternal support, and to my wife Lucia for her love, and decision to share her life with me.

SUMMARY

This thesis incorporates four related studies which examine the ability of school-children (aged 7-11 years) to utilize the psycho-physical concept of perceived exertion (or effort perception) during cycling exercise. Previous research in this area had been promising, but afflicted by notable methodological limitations, such as the use of the adult-specific Rating of Perceived Exertion (RPE) scale as the principal investigative tool, and rather vague forms of data analyses, meant that new research was merited. Accordingly, the present investigations set out to assess both the performance of the recently developed Children's Effort Rating Table (CERT) and clarify the appropriate method of analysis for studies of this kind.

The key outcomes of the current research were that: (i) when asked to express their overall feelings of exertion during incremental cycling exercise, the children's responses with either the CERT or the RPE scale correlated fairly well with objective indicators of physical exertion - heart rate *and* power output. Used in this *estimation* mode, the CERT appeared to be superior. (ii) When asked to manipulate the exercise loading to match specific perceived exertion levels the objective-subjective intensity associations were, whilst significant, less impressive. In this *production* mode, the children seemed not to be able to use the CERT any better than the RPE scale. (iii) Procedural aspects of testing, such as the temporal nature of the exercise protocols, can influence the responses given, and subsequent interpretations. More precisely, the children were better able to use the CERT in production mode when the exercise was discontinuous than continuous. Further work with the CERT amongst children of different ages, involving exercise of different kinds and occurring in different environments, is now encouraged.

INTRODUCTION

Structure of the Thesis

This thesis reports on a series of four empirical studies conducted between 1992 and 1996 into the ability of children to express their feelings, or perceptions, of exercise effort (exertion) during controlled cycle ergometry. The rationale for this research was the emergence of a new child-specific tool for assessing effort perception, the Children's Effort Rating Table (CERT), which was developed as an alternative and more appropriate measure than that which had been used with adults for over twenty years, the Rating of Perceived Exertion (RPE) Scale.

Each study has been written as a scientific paper and has undergone peer-review and subsequent publication. Together with a fifth published article, a critical overview of research into children and effort perception, these papers form the body of this thesis and appear, by-and-large, as they have done in print. For the sake of consistency, the referencing has been standardized throughout in the exact, yet popular style of the American Psychological Association.

It was recognised from the outset that confining the investigations to one mode of exercise (cycle ergometry) and a narrow band of ages (8-11 years) would threaten the external validity, or generalisability, of the outcomes. However, resource limitations and difficulties inherent to research with children made such focus a pragmatic option. Furthermore, owing to the immaturity of knowledge in this domain, it was always intended that the research would remain in the "laboratory" and deal with fundamental concerns. The value of this approach is its potential to direct future, and possibly more applied, inquiry.

The sequence of the four studies represents a developmental process. Initially, a small-scale pilot study was undertaken to gather preliminary data on how well children could use the CERT in two different exercise situations, typically referred to as *estimation* and *production* modes. In its estimation mode, the children were simply required to use the CERT to reflect the degree of perceived strain at three different exercise intensities. In its production mode, they were required to regulate their exercise efforts to match three experimenter-requested CERT levels. In both situations, the children's CERT ratings were found to be well associated with objective markers of effort; heart rate and power output.

Building on this, the second and third investigations were designed to compare the CERT with its predecessor, the ubiquitous Rating of Perceived Exertion (RPE) Scale. Whilst the RPE scale had evolved for use with adults, some researchers had begun to use it unjustly with children. Accordingly, these two studies sought to examine if children could employ the CERT more accurately than the RPE scale. In addition, it was appropriate at this stage to challenge certain practices of data and statistical analyses which have been adopted previously by most researchers in this field, often without adequate justification.

Study four, which addressed an important and previously unconsidered methodological aspect of effort perception studies - the continuous or discontinuous nature of the exercise protocol - commenced with the focus specifically on the CERT scale being used in its production mode. It was anticipated that the introduction of rest periods between exercise bouts would have a bearing on the children's ability to adjust their cycling resistances in accordance with their sense of effort. This study also considered the influence of the temporal nature of the protocols, that is, the timing of the recording of the dependent variable (heart rate). If this parameter is not observed in a 'steady-state', its association with effort perception ratings could be misleading.

The final section of this thesis (Appendices apart) concludes the current programme of research. Consequently, it summarises the insight acquired and offers important guidance for future investigators in this area.

An Overview of Research on Effort Perception

The extent and significance of research into adults' perceptions of exercise effort (or perceived exertion - the terms are used synonymously) is exemplified in a recently published book co-authored by Noble & Robertson (1996). Simply entitled *Perceived Exertion,* the 12 chapter, 320 page volume chronicles the emergence in the 1960s of this renowned concept from the discipline of psychophysics to its current applications in the sporting, exercise, and clinical fields. The authors, themselves key researchers in this field, have produced a bibliography of 450 publications spanning 33 years of study (1960-1993), which they classify as all having dealt with the exercise perception process *per se* (p. xiv). Many of these articles are cited in their comprehensive synthesis of the perceived exertion research, making the publisher's claim that the book will establish itself as a "standard reference in the field for years to come" seemingly well founded. Accordingly, whilst the remainder of this overview is written to "set the scene" for the ensuing programme of research on effort perception in children, no claim is being made for its completeness.

Noble & Robertson (1996; pp. 43-57) describe in detail how the science of human perception, *psychophysics*, became applied to the study of perceived exertion during exercise, and, in essence, they build up a tenable case for the use of alpha-numeric scales, like the famous Borg 6-20 Rating of Perceived Exertion (RPE) Scale, as a valid means of measuring such subjectivity. Implicit in this argument is that humans experience sensory stimulation (intrinsic and extrinsic) through exercise which they can interpret (perceive) and "report" in a variable and quantifiable way. This resulting 'sense of effort' is that which has been investigated in many different exercise settings over the past 37 years, even though it is argued that evidence has never been produced to support its existence (Noble & Robertson, 1996; p. 45). This paradox can probably be explained by the intrinsic appeal of the notion that if humans do experience effort during exercise and are able to distinguish between different amounts of such effort (stimuli), then they might be able to use this sensitivity to adjust their exercise output. The application of this ability is now established amongst exercise and sports scientists, as well as clinicians, but was first recognised by Gunnar Borg in 1970, who, once a valid and reliable measuring scale had been developed, saw that self-reports of exertion had a place in training and rehabilitation situations. Instead of relying on traditional objective indicators of effort, such as heart rate, power output or speed, to guide exercise intensity, Borg advocated the use of subjective feelings of effort, represented by his initial 15-point RPE scale (Borg, 1970; see Figure 1).

It would not be out of place here to discuss the development of the 'all-important' measurement tool, especially considering the aforementioned concerns over the reality of an effort sense, and, that one might justifiably wonder how we can attempt to measure something that we're not certain exists? However, this will not be done. Instead, by accepting that a tool like the RPE scale has *face* validity (with regard to the verbal expressions used) and to some extent *criterion* validity (with regard to its associations with objective markers of effort), a platform is created from which alternative, subject-specific scales, such as the one featured in the studies of children that follow, can readily be devised and their utility assessed. Moreover, whilst arguments may exist about fundamental aspects of the RPE scale, such as whether it can truly measure individual differences in effort perception, or whether it is ordinal or interval in nature (Noble & Robertson, 1996; pp. 60-69), its undoubted practical appeal (primarily from a physiological perspective) and the lack of a viable alternative has thus far made it a resilient instrument.

The assessment of perceived exertion has become an important component in the clinical and laboratory practice of (i) assessing exercise tolerance, (ii) prescribing exercise intensity (for training), (iii) assessing the effectiveness of a therapeutic exercise intervention, and (iv) guiding the duration of a graded exercise test (Noble & Robertson, 1996; p. 93) This status would not exist had many investigations not established moderate to good associations between effort ratings and numerous physiological and psychological variables or processes, which are thought to provide independent exertion stimuli.

Physiological factors have been traditionally classified as either “central” or “local” (Ekblom & Goldbarg, 1971; Borg & Noble, 1974; Mihevic, 1981), though Noble & Robertson (1996; pp. 105-106) now prefer the terms “respiratory-metabolic” and “peripheral”, and have added a third category, “non-specific”. Respiratory-metabolic processes include measures of cardiac function, such as heart rate and blood pressure, and measures of ventilation, such as rate of expired volume, oxygen consumption, and carbon dioxide production. Peripheral mediators of effort perception are specific to exercising muscles and joints and include factors such as muscle/blood lactate and pH levels, fast/slow-twitch muscle properties, and energy substrate (glucose, free fatty acids, and glycerol) mobilization. These factors are thought to dominate the perception of exertion (Watt & Grove, 1993), particularly muscle lactate accumulation (Pandolf, 1983). Non-specific mediators are considered to include those involved in hormonal and thermoregulation, and those responsible for pain sensation.

In comparison, the influence of psychological factors on perceived exertion has been regarded less highly. Nevertheless, so-called "situational" or environmental factors, such as social influence and expectations of exercise performance, and "dispositional", or personality factors, such as cognitive style, self-efficacy, and stimulus-intensity modulation, have long been considered to account for a third of the variance in perceived exertion (Morgan, 1973), or even more (Watt & Grove, 1993). Indeed, psychological mediators assume a key location in both of Noble & Robertson's new (1996) *Psychophysiological Model of Perceived Exertion* (p. 191) and *Global Explanatory Model of Perceived Exertion* (p. 299), which they promote as "working models" to generate future research, and seemingly represents their attempt to synthesise and prioritise the variety of influential factors.

At this point, it is important from a scientific perspective not to over-emphasise the significance of the determinants of effort perception. It is true that research in this domain has proceeded from the simple premise that the quantification of effort sense with an RPE scale represented a general perception - a so-called 'Gestalt' (formed from the input of many sensory signals and their interpretation), to a multifarious, though not necessarily systematic, exploration into the influences of individual physiological and psychological sensory cues. However, a dearth of experimental studies has not enabled the principal causal elements of perceived exertion to be identified. It is rather surprising, therefore, that this shortcoming does not appear to have thwarted exercise scientists from demonstrating the utility of the effort perception concept; over 200 papers from Noble & Robertson's (1996) bibliography deal with applications of RPE. Perhaps a Gestalt understanding of perceived exertion is sufficient after all?

So, how has the concept of effort perception actually been applied with such regularity? Most, if not all investigations prior to 1980 used Borg's (1970; Figure 1) RPE scale in evaluating the subjective strain experienced during dynamic exercise, often at specific power outputs during graded exercise tests (Noble, 1982). Subjects would simply report (or point to) a number on the scale which best indicated their present feeling of exertion. This use of RPE has come to be known as its 'response' or 'estimation' mode (Myles & Maclean, 1986; Eston & Williams, 1988; Dunbar, Robertson, Baun, Blandin, Metz, Burdett, & Goss, 1992) and initially formed the basis of Borg's early validation of the scale against heart rate responses (Borg & Linderholm, 1967; Borg, 1970). The second application of RPE, its 'production' (or regulation) mode, was first tested empirically by Smutok, Skrinar, & Pandolf (1980), though (as mentioned above) Borg had advocated the principle 10 years earlier. This involved subjects using their understanding of the perceived exertion concept to help them regulate their exercise output. In effect, subjects were asked to utilize their effort sense to produce (select) exercise intensities that 'matched' experimenter-identified RPE levels, such as 11, 13, and 17.

The choice of exercise mode has varied (cycling, walking, running, rowing, swimming, and stepping), but the protocols have tended to be multi-stage (each 2-4 minutes), changeable (incremental or randomised loadings) and

Figure 1. Rating of Perceived Exertion (RPE) Scale (Borg, 1970)

6	
7	Very, very light
8	
9	Very light
10	
11	Fairly light
12	
13	Somewhat hard
14	
15	Hard
16	
17	Very hard
18	
19	Very, very hard
20	

progressive (continuous or discontinuous). No standardised procedures have emerged, though most estimation studies have tended to utilise a continuous, incremental (to volitional exhaustion) protocol, quite often as a precursor to a discontinuous, randomised production trial. The laboratory practical described by Williams & Eston (1996) illustrates this custom and shows how data from the estimation trial are extrapolated and used to assess subjects' accuracy of perception during the production trial. That is, the heart rates and power outputs produced at specified RPE levels are compared to those predicted (or observed) from the estimation trial.

In both its modes, such perceptual ability has been typically quantified via correlational analysis of RPE ratings with simultaneously recorded measures of heart rate or oxygen consumption, and/or power output or speed. High positive correlations ($r > 0.90$) have been frequently reported and interpreted as establishing the validity of the RPE scale; changes in objective effort being closely followed by corresponding changes in RPE ratings, and vice versa. However, some researchers have questioned the appropriateness of the statistical technique often used and suggested that such relationships are generally not so strong (Gillach, Sallis, Buono, Patterson & Nader, 1989). In addition, despite claims of being "quite acceptable" (Noble & Robertson, 1996; p. 70), the reliability of effort perceptions has never been adequately established. By definition, such a shortcoming undermines the validity of the concept.

Despite these basic deficiencies, research activity related to effort perception has been maintained, showing a bias towards the use of RPE in its production mode (Glass, Knowlton, & Becque, 1992; Koltyn & Morgan, 1992; Dunbar, Goris, Michielli, & Kalinski, 1994; Parfitt, Eston, & Connolly, 1996; Shephard, Kavanagh, Mertens, & Yacoub, 1996). Also, as the concept continues to permeate new sports and exercises (both recreational and occupational), such as aerobics (Clapp & Little, 1994), swimming (Ueda & Kurokawa, 1995), rowing, (Marriott & Lamb, 1996), stepping (Walker, Lamb & Marriott, 1996) and wood cutting (Hagen, Vik, Myhr, Opsahl & Harms-Ringdahl, 1993), its validity in estimation mode continues to be examined. Evidently, the popularity and unquestionable status of RPE in the adult population shows little sign of waning. However, this situation does not apply to children.

Purposefully omitted from the above 'developments' is reference to the fairly recent expansion of interest in effort perception by paediatric exercise scientists and physical educators. Previously, apart from original work by Bar-Or (1977) and a somewhat prophetic discussion paper by Eston (1984), only on rare occasions had investigators connected RPE with children. But, between 1989 and 1993, nine published articles dealt intentionally with the abilities of small groups of children (aged 8-17) to apply the 6-20 RPE scale during controlled exercise situations. As with adults, the appeal of the perceived exertion notion was its potential for allowing exercise intensity to be governed by an individual's effort sense. For children, this utility was, and remains, particularly relevant to the delivery of physical education and health promotion.

Whilst some 'success' was claimed from these studies, for example, RPE used in estimation mode correlating with heart rate (r = 0.64 - 0.87) and oxygen uptake (r = 0.84), it became apparent to certain researchers that performance was highly variable. Moreover, many children, especially those younger than 10 years, found great difficulty relating to the RPE scale (applied in either of its modes). Consequently, and four years after a recommendation by Bar-Or & Ward (1989), an effort perception scale for children, the Children's Effort Rating

Table (CERT), was devised and pilot tested by Williams, Furlong, Hockley & Mackintosh (1993) amongst a sample of 4-9 year-olds, and the results reported in detail by Williams, Eston & Furlong (1994). The empirical investigations that follow explore in more detail the efficacy of this new tool and represent the current level of understanding of children's perceptions of exercise effort.

CRITIQUE OF LITERATURE

Effort Perception in Children[1]

[1]The content of this paper has appeared in a 'Leading Article' by Lamb, K.L. & Eston, R.G. (1997) *Sports Medicine*, 23, 139-148.

Introduction

The interest in adults' perceptions of exercise effort has been buoyant for over 30 years and remains the target of considerable research activity amongst exercise and sport scientists. Readers of the journal *Sports Medicine* will be familiar with three reviews of literature which comprehensively address such activity (Carton & Rhodes, 1985; Watt & Grove, 1993; Williams & Eston, 1989). The practice of using an alpha-numeric scale, particularly the two versions of the ubiquitous 6-20 Rating of Perceived Exertion (RPE) scales (Borg, 1970; Borg, 1985) as a marker of subjective effort is common place in a variety of exercise environments. No longer the domain of the exercise physiology laboratory or cardiac rehabilitation unit, the utility of the more recent RPE scale (Figure 2) has been extended into the community as providers of exercise have recognised its worth amongst heterogeneous groups of participants. Being a relative concept, RPE can help in the delivery of 'appropriate' levels of exercise, from the swimming pool or aerobics studio, to the fitness suite treadmill. It is rather surprising, therefore, that providers of exercise to *children* have yet to realize, or be convinced, that such a concept might also be employed usefully in their domain.

The following synopsis will show that, despite its present lack of use, a body of research knowledge on children and effort perception has developed over the last 20 years. In evaluating such information, the intention is to illuminate the likely reasons for the current status, and offer suggestions to guide future research in this area.

Historical Perspective

The pioneer of research into children's perceptions of exercise effort was Oded Bar-Or, who in 1975 presented data collected from six different projects, involving a total of 589 children (aged 7-17 years) to the First International Symposium on Physical Work and Effort, in Stockholm. Published in 1977 (Bar-Or, 1977) the data represent the children's RPEs recorded during continuous, incremental cycle ergometry, over an exercise range of 50-200 watts. By plotting RPE against heart rate, Bar-Or revealed that children in each of six discrete age-groups (7-9, 8-10, 10-11, 11-12, 13-14, and 16-17 years) reported higher RPEs in line with increases in the cycling resistance. Moreover, it was concluded that, with one exception (the 7-9 year-old group), these children gave lower ratings of effort than adults at the same relative exercise intensity.

Whilst the above research contained a number of inconsistencies, such as the exercise habits of the children being variable, and some of the data being collected as part of a heat-acclimatization study, it acquired a near-definitive status for the next ten years. With a few exceptions, notably a somewhat perplexing abstract reporting on the consistency of RPE ratings among 'active' girls (Kahle, Ulmer & Rummel, 1977) and a study examining perceived exertion levels during 60 minutes continuous, constant-load cycling among predominantly

Figure 2.
The 6-20 RPE Scale
(Borg, 1985)

6	NO EXERTION AT ALL
7	
8	EXTREMELY LIGHT
9	
10	VERY LIGHT
11	LIGHT
12	
13	SOMEWHAT HARD
14	
15	HARD (HEAVY)
16	
17	VERY HARD
18	
19	EXTREMELY HARD
20	MAXIMAL EXERTION

Figure 3.
The CERT scale
(Williams, Eston & Furlong, 1994)

1	VERY, VERY EASY
2	VERY EASY
3	EASY
4	JUST FEELING A STRAIN
5	STARTING TO GET HARD
6	GETTING QUITE HARD
7	HARD
8	VERY HARD
9	VERY, VERY HARD
10	SO HARD I'M GOING TO STOP

female anorexic adolescents (Davies, Fohlin & Thoren, 1980) no further reports relating to children appeared in the academic literature until the middle of the next decade. After this time, a slow, but regular succession of reports started to emerge from North America, Great Britain, and Asia (Bar-Or & Reed, 1986; Eston & Williams, 1986; Miyashita, Onedera & Tabata, 1986; Van Huss, Stephens, Vogel, Anderson, Kurowski, Janes, & Fitzgerald, 1986; Ward, Blimkie and Bar-Or, 1986).

Researchers continued to examine the RPE-objective effort relationship primarily in the laboratory setting and in the so-called *estimation* mode, which simply requires that a subject chooses an RPE value appropriate to a given exercise intensity. Most studies have chosen a cycle ergometer as the exercise medium (Alekseev, 1989; Gillach *et al*, 1989; Meyer, Bar-Or & Wilk, 1995; Ward & Bar-Or, 1990; Ward, Jackman & Galiano, 1991) whilst others have used a motorised treadmill (Eakin, Finta, Serwer, & Beekman, 1992; Mahon & Marsh, 1992; Mahon & Ray, 1995; Tolfrey & Mitchell, 1996) an arm ergometer (Ward, Bar-Or, Longmuir, & Smith, 1995) or a swimming pool (Ueda and Kurokawa, 1991). Only two studies (Nystad, Oseid & Mellbye, 1989; Stratton & Armstrong, 1994) have used the RPE scale in the 'field' setting of a school physical education lesson in order to explore whether children's effort perceptions were related to their measured heart rates.

Since 1990, seven articles have been published which deal with the ability of children to use their perceptions of effort to regulate exercise intensity. This *production* mode of perceived exertion traditionally requires subjects to adjust their exercise load (generally speed or power output) to match certain experimenter-specified RPE values. Again, the cycle ergometer has predominated (Ward & Bar-Or, 1990; Ward *et al*, 1991; Williams, Eston & Stretch, 1991), though three investigations have ventured out of the laboratory onto a running track (Ward & Bar-Or, 1990; Ward *et al*, 1991; Ward *et al*, 1995) in an attempt to confront the external validity of the RPE scale.

A significant development in the study of children's perceptions of exercise effort occurred in 1993/1994 with the publication of two papers concerned with the validity of a new child-specific rating scale (Williams *et al*, 1993; Williams *et al*, 1994, Figure 3). The study that follows this chapter (Study 1) investigates the validity of this aptly named Children's Effort Rating Table (CERT) during cycling exercise. Prior to the appearance of the CERT, little practical regard had been given to the suitability of the RPE scale for children. Exempt from this criticism is a report provided by Nystad *et al* (1989) who attempted to clarify the RPE's intensity dimension by replacing its nine descriptors ("very hard", "hard", "light" and so on) with six stick-figures depicting various stages of fatigue (see Figure 4).

Whilst time will reveal whether CERT is actually superior to RPE, it was at least devised with children in mind. In this respect, it has five fewer possible responses, a range of numbers (1-10) more familiar to children (than 6-20), and verbal expressions that not only accompany all 10 numbers, but represent words chosen and understood by children as descriptors of exercise effort (Williams *et al,* 1994). Progress in this challenge to an 'establishment' is addressed in Studies 2 and 3 of this thesis, which examine the validity of both scales during *estimation* and *production* modes.

The Validity and Reliability of Children's Effort Perceptions

These fundamental characteristics of any assessment tool should, and often do, receive much attention from scientists before the tool becomes an accepted item for use in future investigations. With regard to the available options for assessing children's perceptions of effort, it cannot be argued convincingly that they possess either admissible validity or reliability. In the case of the latter, there has been little regard at all. There remains ample scope for these concerns to be addressed.

Of course, any consideration of validity and reliability amongst children should be accompanied by a regard for their age, or more precisely, their ability to read or understand the chosen scale. Indeed, it was recognised at the aforementioned 1975 conference that perception of exercise effort is a developmental issue (Borg, 1977) which might strongly be influenced by the extent of children's experiences of exercise. Surprisingly, no research has incorporated this key aspect into its design. However, it does now seem to have been recognised that Borg's RPE scale is probably unsuitable for some children, particularly young children, and that the employment of alternative scales is justified (Williams *et al*, 1991; Williams *et al*, 1994; Noble & Robertson, 1996, p. 301))

Estimation Mode

Typically, and in line with the kind of research previously conducted on adults, children's perceptions of effort during estimation mode have been validated against accepted objective measures of physiological strain, such as heart rate (HR), power output, or oxygen uptake (Bar-Or, 1977; Bar-Or & Reed, 1986; Eston & Williams, 1986; Miyashita *et al*, 1986; Ward, Blimkie & Bar-Or, 1986; Alekseev, 1989; Gillach *et al*, 1989; Nystad *et al*, 1989; Ueda & Kurokawa, 1991; Eakin *et al*, 1992; Williams *et al*, 1994; Duncan, Mahon, Gay & Sherwood, 1996). However, the degree of validity reported, usually expressed as a bivariate (Pearson) correlation (*r*), has not only been variable (from 0.45-0.99), but in some cases questionable in terms of the type of analysis adopted by researchers (see below).

Bar-Or's early validation of the RPE scale with groups of children aged 7-17 years was initially encouraging as correlations as high as 0.88 were observed for the oldest (16-17 year-old) group (see Table 1). Subsequent studies of groups of healthy children aged 10 years and above consistently produced validity correlations of 0.75-0.86 during cycle ergometry (Bar-Or & Reed, 1986; Eston & Williams, 1986; Ward *et al*, 1986; Alekseev, 1989; arm cranking, (Bar-Or & Reed, 1986) or treadmill exercise.(Eakin *et al*, 1992). Two studies using a

Table 1.
Exercise Intensity *Estimation* Using Subjective Ratings of Exertion: Exercise Protocols and Validation Analyses with Children

Study	Age (yr)	Protocol	Validity
Bar-Or (1977)	7-17	**Continuous:** *Cycle*, incremental; 3 min bouts: 50-200 W (depending on age).	RPE[1] versus HR; $r = 0.76\text{-}0.88$ (approx.)
Kahle *et al* (1977)	7-11	**Discontinuous:** *Cycle*, incremental; 5 x 1 min bouts @25, 50, 75, 100 & 125% of PWC_{170}, interspersed with 1.5 min rests.	Not addressed.
Davies *et al* (1980)	12-18	**Continuous:** *Cycle*, constant load; 60 min @ 62% VO_2max (RPE reported every 10 min).	Not addressed.
Van Huss *et al* (1986)	8-15	**Discontinuous:** *Treadmill*, incremental; 3 min bouts separated by 3 min rest: 6 mph @0% & 5% grade, then increasing 1 mph & 1% grade until VE.	RPE versus HR; "linear" (girls). RPE versus VO_2; "linear" (boys and girls).
Bar-Or & Reed (1986)	9-19	**Continuous:** *Cycle* or *Arm Cranking*, incremental; 3-6 x 2 min bouts until cadence < 50 rpm.	RPE versus HR (arms); $r = 0.58\text{-}0.81$ RPE versus HR (legs); $r = 0.69\text{-}0.82$ RPE versus %PP (arms); $r = 0.74\text{-}0.92$ RPE versus %PP (legs); $r = 0.72\text{-}0.86$
Eston & Williams (1986)	15-17	**Continuous:** *Cycle*, random; 3-4 min bouts, 30, 60 & 90% of predicted maximal power output.	RPE versus HR; $r = 0.74$, $p < .01$ RPE versus PO; $r = 0.78$, $p < .01$
Miyashita *et al* (1986)	7-18	**Continuous:** *Cycle*, 3-5 four min bouts (further detail not specified).	RPE versus %HRmax; $r = 0.55\text{-}0.94$, $p < .01$

continued

Table 1. (continued)

Study	Age (yr)	Protocol	Validity
Ward *et al* (1986)	14-17	**Continuous**: *Cycle*, incremental (unspecified) to VE.	RPE versus %PP; r = 0.84-0.86 (boys), r = 0.79-0.83 (girls).
Nystad *et al* (1989)	10-12	**Continuous**: *Physical Education Lesson*, 10 x 1-hour lessons.	RPE versus HR; r = 0.05
Gillach *et al* (1989)	10-14	**Continuous**: *Cycle*, incremental; 25 W/2 min to 85% predicted maximum heart rate.	RPE versus HR; r = 0.64-0.65, $p < .01$
Alekseev (1989)	10-14	**Continuous**: *Cycle*, incremental (unspecified) to VE.	RPE versus HR; r = 0.84, $p < .01$ RPE versus HRmax; r = 0.87, $p < .01$
Ward & Bar-Or (1990)	9-15	**Continuous**: *Cycle*, incremental & random; bouts of 20, 40, 60 & 80% peak aerobic power every 2 min.	Not addressed.
Ward *et al* (1991)	8-14	**Continuous**: *Cycle*, incremental; 25 W/2 min to VE.	Not addressed.
Ueda & Kurokawa (1991)	10-12	**Discontinuous**: *Swim*, tethered, incremental; 1 kg/5 min (separated by 10-20 min rest).	RPE versus %VO_2max; r = 0.816-0.997 (for individual subjects).
Eakin *et al* (1992)	10-17	**Continuous**: *Treadmill*, incremental; variable speed, 3-min stages, 2% slope increase (from 10%) to VE.	RPE versus HR; r = 0.83-0.87, $p < .01$ RPE versus VO_2; r = 0.84-0.85, $p < .01$
Mahon & Marsh (1992)	8-12	**Continuous**: *Treadmill*, incremental; fixed speed, 2% slope increase/min to 'VO_2max'.	Not addressed.
Stratton & Armstrong (1994)	12-13	**Continuous**: *Physical Education (Handball) Lesson* (unspecified).	RPE versus HR; analysis unclear.

continued

Table 1. (continued)

Study	Age (yr)	Protocol	Validity
Williams *et al* (1994)	4-9	**Continuous:** *Bench Step*[2], incremental; 4 x 2-min bouts with loadings of 0, 5, 10 & 20% body mass.	CERT[3] versus HR; r = 0.73-0.99, $p < .01$
Meyer *et al* (1995)	9-12	**Discontinuous:** *Cycle*, incremental; 3 x 15 min @50% peak VO_2, then @90% to VE (separated by 10 min rest).	Not addressed.
Mahon & Ray (1995)	9-11	**Continuous:** *Treadmill*, incremental; fixed speed, 2% slope increase/min to VE.	Not addressed.
Ward *et al* (1995)	11-30	**Continuous:** *Arm ergometry*, incremental; 3-6 x 2 min loadings (unspecified) to VE.	Not addressed.
Duncan *et al* (1996)	9-11	**Continuous:** *Cycle*, incremental; 10 W/1 min to VE. *Treadmill*, incremental; 3 mph, 2.5% slope increase/min to VE.	RPE versus HR; mean r = 0.98 RPE versus HR; mean r = 0.98
Tolfrey & Mitchell (1996)	11-14	**Discontinuous:** *Treadmill*, incremental; 3-min stages (unspecified) to VE.	Not addressed.

[1]Rating of Perceived Exertion. [2]Information unclear, but assumed continuous. [3]Children's Effort Rating Table.
VE = Volitional Exhaustion; HR = Heart Rate (b/min); PWC_{170} = Physical Work Capacity at 170 b/min; HRmax = maximum (age-related) Heart Rate; PO = Power Output (watts); PP = Peak Power (watts); VO_2 = Oxygen Uptake (ml/kg/min); VO_2max = maximal Oxygen Uptake (ml/kg/min).

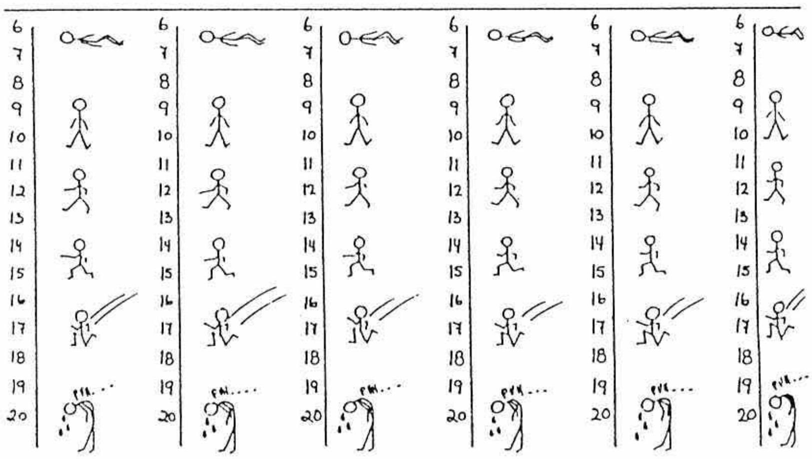
Japanese version of the RPE scale during cycling (Miyashita *et al*, 1986) and tethered swimming (Ueda & Kurokawa, 1991) reported correlations in excess of 0.90 and 0.82-0.99, respectively, though it was apparent that these values reflected the data of individuals, and not groups.

Research involving younger children has been less common, possibly because of concerns expressed regarding their ability to understand the RPE scale. Of the groups of children studied by Bar-Or (1977), it was observed that the youngest (aged 7-9 years) were the least 'accurate' when estimating their exercise effort. Moreover, Miyashita *et al's* (1986) findings that the RPE-HR correlations were the lowest for those aged 7-9 years (r = 0.55-0.74) lead to the assertion that the critical age for understanding the Japanese RPE scale was 9 years. Despite this evidence of poor validity most of the evidence for the inappropriateness of the RPE scale with young children has been anecdotal. Whilst it is probably true that they do have difficulty relating to its format, not least the meaning of some of its words, no scientific study has yet set out to address this issue.

On this theme, it is rather unfortunate in several respects that little attention has been paid to the innovative study conducted by Nystad *et al* (1989) Not only were these researchers the first to consider the external validity of the RPE scale by testing its application in the real-world setting of a 60-minute physical education class, but they also endeavoured to enhance their subjects' understanding of the scale by using drawings of stick-figures depicting varying degrees of effort (see Figure 4). No other studies have used RPE in this way.

The lack of an RPE-HR association observed in the above study was explained in terms of the small sample size ($n = 10$), the relatively low intensity nature of the physical education lesson (heart rates mostly below 150 b/min), and importantly, the fact that being asthmatic (to varying degrees) caused the children

Figure 4.
Borg Scale with Stick Figures
(Nystad *et al,* 1989)



to report disproportionately higher RPEs than their heart rates. Whilst these limitations were undoubtedly influential, Nystad *et al* also highlighted another confounding factor whose importance has still yet to be fully appreciated. Their argument was that their subjects (aged 10-12 years) may have been physically inexperienced (lacking a consciousness of different exercise intensities), and, more serious, not really able to understand the concept of perceived exertion when it was explained to them.

The formation of the CERT scale was a genuine attempt to tackle this problem of children's concepts of exercise effort. The initial validation of the CERT using bench-stepping as the mode of exercise (Williams *et al*, 1994) produced an almost perfect correlation (r = 0.99) between exertion level and heart rate amongst 8 and 9 year-olds, and a lower, yet respectable association for 6-7 year-olds (r = 0.73). However, it was acknowledged that owing to the low and narrow range of exercise intensities administered (149-171 b/min), these correlations may have been spuriously high. This issue is clarified in Studies 1 and 2 of this thesis.

In studies of adults, the *reliability* of the RPE scale (in estimation mode) has been seemingly settled on the evidence presented in a couple of reports published in the 1970s (Skinner, Hutsler, & Buskirk, 1973; Stamford, 1976). Whilst one could question this claim for reliability on statistical grounds (see the section below), the issue of reliability in children is far less clear. Moreover, current and future researchers are advised to assess for themselves the reliability of the RPE, or any alternative effort perception scale.

The few published studies to date have produced varying accounts of reliability. Ward & Bar-Or (1987) reported a 2-day test-retest correlation of 0.86 for the RPE scale during cycling with a group of 20 obese children (aged 9-15 years), and elsewhere on the same sample (Bar-Or & Ward, 1989) correlations (presumably Pearson product-moment) ranging from 0.59-0.92 for specific exercise intensities, with the higher coefficients pertaining to the higher exercise intensities. More recently, Mahon & Marsh (1992) reported a 2-16 day test-retest Pearson correlation of 0.78 for RPEs recorded at the ventilatory thresholds of 30 children (aged 8-12 years), measured during treadmill exercise.

Production Mode

Where perceptions of effort during production mode have been validated, researchers have tended to compare the objective indicators with *expected* values calculated via regression analysis of individuals' data from a previous estimation trial (Ward & Bar-Or, 1990; Ward *et al*, 1991; Ward *et al*, 1995). In this way, values of heart rate, power output, or oxygen uptake corresponding to each number on the effort perception scale can be predicted, and then compared to the values achieved in the production trial. This comparison has then been quantified either by analysis of variance techniques (Ward & Bar-Or, 1990; Ward *et al*, 1991; Ward *et al*, 1995) or interclass correlation (Williams *et al*, 1994; see Table 2). In effect, the ability of children to use perceptions of effort to self-regulate exercise output has been traditionally compared to approximations of their ability to rate or estimate the exercise intensity of given work bouts.

However, the validity of children's use of RPE in production mode is not clear. For example, in the first full paper published on this theme (Ward & Bar-

Table 2.

Exercise Intensity *Production* Using Subjective Ratings of Exertion: Exercise Protocols and Validation Analyses with Children.

Study	Age (yr)	Protocol	Validity
Ward and Bar-Or (1990)	9-15	**Discontinuous:** *Cycle*, random; 4 min @RPE[1] 7, 10, 13, & 16 separated by > 3 min rest. **Discontinuous:** *Track walk/run*, random; 1x400 m lap @RPE 7, 10, 13, & 16 separated by > 3 min rest.	*Cycle*: HR higher ($p < .05$) at each RPE level (RPE16 > RPE13 > RPE10 > RPE7). *Track*: HR @RPE 7 lower ($p < .05$) than at other RPE levels.
Ward *et al*, (1991)	8-14	As above.	*Cycle*: %PP higher ($p < .05$) at each RPE level RPE16 > RPE13 > RPE10 > RPE7). *Track*: %PP @RPE 7 lower ($p < .05$) than at other RPE levels.
Williams *et al* (1991)	11-14	**Continuous:** *Cycle*, incremental; 4 min @RPE 9, 13 & 17.	Significant ($p < .01$) effect of RPE level on HR.
Williams *et al* (1994)	4-9	**Continuous:** *Bench-step*[1], random; 1 min @CERT[2] 5 & 7.	Non-significant ($p > .05$) correlation between HR_{pred} and HR_{obs}.
Ward *et al*, (1995)	11-30	**Discontinuous:** *Track*[3], random; 1x400 m lap @RPE 10, 13, & 16, separated by rest until HR < 100 b/min. .	HR higher ($p < .01$) @ each RPE level (RPE16 > RPE13 > RPE10 > RPE7). Wheeling speeds higher ($p < .05$) @ each RPE level (RPE16 > RPE13 > RPE10 > RPE7).

[1]Rating of Perceived Exertion. [2]Children's Effort Rating Table. [3]Subjects were wheel-chair bound.

HR = Heart Rate (b/min); HR_{pred} = Heart Rate predicted from regression analysis; HR_{obs} = Heart Rate observed; PO_{pred} = Power Output predicted from regression analysis; PO = Power Output (watts); PP = Peak Power (watts); PO_{obs} = Power Output observed.

Or, 1990), it was concluded that whilst the sample of overweight children (aged 9-15 years) produced exercise intensities during cycling that were significantly different to expected (or "criterion") values, they *were* able to discriminate between each of the four prescribed RPE values (7, 10, 13, and 16). Very similar findings were reported from an almost identical study one year later (Ward *et al*, 1991) and from a subsequent study of wheelchair-bound subjects (aged 11-30 years) engaged in wheeling around a track (Ward *et al*, 1995). Moreover, whilst the study of Williams *et al* (1991) indicated that children (aged 11-14 years) were able to increase exercise output (reflected by HR) to match increases in effort perceptions (RPEs), a later study (Williams *et al*, 1994) found non-significant correlations between predicted and actual HRs produced at two levels of the CERT scale.

An indication of production *reliability* can be gleaned from a couple of studies using the RPE scale (Williams, Eston & Stretch, 1991; Ward *et al*, 1995). Indirect evaluation of reliability was provided by Williams *et al* (1991) who reported that the mean heart rates produced by 11-14 year-old children at levels 9, 13, and 17 were not significantly different over the course of three identical cycling trials. Furthermore, Ward *et al* (1995) claimed (albeit without supporting analysis), that their sample of wheelchair-bound children and adults "displayed an excellent retention" over one month of an ability to regulate wheeling intensities using RPE.

Statistical Concerns

The above review of the validity and reliability of children's perceptions of effort is clouded by the apparent mis-use of certain statistical techniques by the majority of the researchers cited. This practice is commonplace in the exercise and sports sciences (Nevill, 1996; Bartlett, 1997). probably due to the habit some researchers have of simply copying the statistical methods used in previous related publications. To reiterate, knowledge of validity and reliability is fundamental to the use of a particular measurement tool. Therefore, particular attention to these concerns is justified in advance of the investigations which follow this chapter.

Gillach *et al* (1989) questioned the degree of validity attributed to the RPE scale when used in estimation mode on the grounds that the bivariate (Pearson) correlations (typically used to reflect validity) were seemingly calculated incorrectly. They argued that researchers' habit of calculating correlation coefficients for each subject in a sample, and then determining a mean value from these (see Duncan *et al*, 1996) was wholly inappropriate and likely to inflate the magnitude of the relationship between perceptions of effort and objective measures. This problem arises because correlation coefficients are susceptible to the number of data pairs being analyzed and the heterogeneity of these data; those coefficients determined from four or five data pairs can be 'high' simply due to the influence of one or two 'extreme' values (Edwards, 1976, pp. 55-56), and not because a genuine relationship exists. In Gillach *et al*'s study, the preferred option of calculating coefficients from all subjects' data simultaneously had the effect of reducing the *r*s from 0.95 to 0.64 (in adults) and 0.94 to 0.65 (in children), making the validity of effort perceptions look markedly less impressive.

With regard to the validity, or accuracy, of effort ratings during production mode, previous studies have tackled this by: (i) examining the consistency of the objective effort-perceived effort relationship across estimation and production trials (Ward & Bar-Or, 1990; Ward *et al*, 1991; Ward *et al*, 1995), and (ii) comparing, via analysis of variance (Ward & Bar-Or, 1990) or correlation (Williams *et al*, 1994), the produced exercise intensities (heart rate and power output) to 'criterion' values calculated from an estimation trial. Until now, the appropriateness of such statistical techniques has gone unquestioned.

Studies that have adopted method (i) have determined individual (subject-by-subject) linear regression equations for each trial in order to produce group means of *slope* and *intercept* values. Analysis of variance (ANOVA) has then been employed to compare these regression characteristics across trials[1]. This type of analysis seems unsuitable on two accounts. Firstly, as described above, the forming of *individual* regression equations from a small number of data pairs can make relationships between variables appear spuriously strong. The regression characteristics (r^2, slope and intercept) would therefore mis-represent

---

[1]This method should not be confused with the analysis of the *homogeneity* of the regression slopes and intercepts, described by Kerlinger & Pedazhur (1973; ch. 10), which relates to data collected on independent (and not dependent) groups.

the bivariate relationship for each subject, as would the mean characteristics then calculated to represent the bivariate relationship of the whole group. Secondly, and this criticism also applies to method (ii), it is difficult to accept that the information gathered from an estimation trial, in which the exercise has typically been continuous and the ratings of effort on-going, is directly comparable to that gathered in a production trial, in which the exercise has typically been discontinuous (with recovery periods) and the ratings of effort have been retrieved from memory. Indeed, Noble's (1982) argument that two dissimilar psychophysical processes are involved here is not only pertinent, but for children is probably confounded by the extent of their perceptual development.

On account of this doubt, and in the absence of a convincing rationale for continuing to adopt either or both of the 'established' methods, it seems reasonable that future studies concerned with the ability of children to use an effort rating scale to guide their exercise responses, should focus simply on the data generated in a production trial. Accordingly, repeated measures ANOVA could then be used to assess the variability of performances. Studies 3 and 4 of this thesis represent the first to adopt such an approach.

The few claims for the reliability of effort perceptions (during either estimation or production mode) made in the studies referred to above are unfounded because of their reliance on the interclass correlation coefficient. A strong case has been argued in recent years that the more appropriate statistic for univariate test-retest reliability analysis is the intraclass coefficient (*R*). This is obtained from a repeated measures analysis of variance (Safrit & Wood, 1989; Thomas & Nelson, 1990; Vincent, 1994) and calculated as R = (MSs - MSw)/ MSs, where MSs represents *between-subject* variance, and MSw represents *within-subject* variance. The *interclass* Pearson correlation is a bivariate statistic which measures the degree of association between two independent variables, and not the degree of agreement between repeated measurements of the same variable. Consequently, test-retest scores can be quite (significantly) different, yet remain highly correlated because they have changed in a systematic way. In addition, the interclass correlation cannot cope with more than two sets of data at once; a situation which does not always hold in studies of reliability, where the consistency of performance over several trials is often of interest to researchers. These limitations do not apply to the intraclass correlation (Vincent, 1994, p. 178).

Methodological Considerations

It is striking from Table 1 that most investigations of effort perception among children (in estimation mode) have adopted one particular type of exercise protocol; a continuous, relatively long (up to 20 min), incremental test on a cycle ergometer. Whilst the consistent choice of exercise medium is excusable on practical grounds, it is surprising that a more varied selection of exercise protocols has not been utilised. As a consequence, much of our understanding of children's capability to rate their exercise effort has evolved from measuring their responses to a situation in which they are aware (or quickly realize) that the exercise is getting progressively more demanding. One might attribute the apparent validity of children's perceptions to this factor alone.

To date, no study has investigated systematically the influence of exercise protocol on children's perceptions of physiological effort. Furthermore, only two have randomised the order of presentation of work loads (Eston & Williams, 1986; Ward & Bar-Or, 1990), though one of these was designed initially to compare random and incremental protocols (Ward & Bar-Or, 1990). Unfortunately, little of note emerged from this comparison, other than it being reported that "no differences were observed". Studies which have allowed rest periods between exercise bouts (Kahle *et al*, 1977; Van Huss *et al*, 1986; Ueda & Kurokawa, 1991; Meyer *et al*, 1995; Tolfrey & Mitchell, 1996), and thus potentially lessened the influence of fatigue on effort perceptions, have still been incremental in nature.

The importance of this inconsistency or inappropriateness of methodology is heightened where researchers have tried to apply the concept of effort perception in production mode. On reflection, the practice of comparing exercise intensities generated at particular perceived effort levels during discontinuous, randomised protocols with those estimated from continuous, incremental estimation protocols, seems illogical. To reiterate earlier comments, a reasonable, and simple strategy would be to focus attention solely on examining how children perform during the production trials; how they respond during estimation trials is not important. If it can be firmly established, and some evidence already exists (Ward & Bar-Or, 1990; Ward *et al*, 1991; Williams *et al*, 1991; Ward *et al*, 1995), that children are able to adjust their physical exertion in proportion to their perception of effort, then there is genuine potential for applying this relationship into the physical education/health promotion environment, as alluded to over a decade ago (Eston, 1984).

It has been suggested several times that studies of perceived exertion in children deserve to be administered differently to those in adults (Bar-Or, 1989); Bar-Or & Ward, 1989; Williams *et al*, 1991; Williams *et al*, 1994). However, despite the modifications made to the RPE scale by Nystad *et al* (1989) and the development of the child-specific CERT - attempts to overcome comprehension difficulties - surprisingly few studies have set out to unearth the variables which impact upon a child's perception of exercise effort. Variables such as habituation, sex, activity level, reading ability, exercise mode, or level of perceptual development are possible influences which have not been investigated. Yet, it could be argued that such sophisticated research is superfluous if all that is required is a generic tool which allows children to appreciate the breadth of the physical exertion continuum, and be able to adjust their own efforts on the basis of such consciousness.

Conclusions

Although research into children's perceptions of physical effort has a history, our present level of understanding remains rather limited. Investigators can be excused initially perhaps for having conducted their research in the same vein as that performed in greater volume on adults, but the time for progress is now overdue. A lack of consensus currently exists in terms of how data should be gathered and analysed, making interpretations of validity and reliability quite difficult. Future investigations with children should keep apart the two traditional applications of effort perception, estimation and production, and endeavour to agree on suitable research designs. These designs should account for factors related to the nature of the exercise protocols, such as whether they are continuous or discontinuous, or whether the dependent variables should be measured after two or five minutes of exercise. In general, more attention deserves to be given to assessing the efficacy of alternatives to the RPE scale, like the CERT, in order that prospective consumers of this knowledge, such as physical educators/health promoters, can begin to contemplate its usefulness beyond the laboratory.

STUDY 1

Validity of a Perceived Exertion Scale for Children[2]

[2] The content of this chapter has appeared in the paper by Eston, R.G., Lamb, K.L., Bain, A., Williams, A.M. & Williams, J.G. (1994). *Perceptual and Motor Skills*, 78, 691-697.

Introduction

The use of rating scales to assess perceived levels of exertion, particularly the 6 to 20 Rating of Perceived Exertion (RPE) scales developed by Borg (1970; 1985), has been established as a valid method of estimating actual physiological demand during exercise by adults, and also as a tool for monitoring and regulating exercise intensity during structured training programmes (American College of Sports Medicine, 1991; Dunbar *et al*, 1992; Eston, Davies, & Williams, 1987; Eston & Williams, 1988; Watt & Grove, 1993). Few studies have considered whether RPE scales can be employed to regulate levels of physical activity among children or adolescents, for whom a potential application is in the delivery of optimal (cardio-respiratory) training intensities during school-based physical education.

Studies of children have generally quantified exercise effort by monitoring heart rate responses to experimenter-determined loadings on a cycle ergometer (e.g., Eston & Williams, 1988; Ward *et al*, 1991; Williams *et al*, 1991). In some research subjects were required to control (regulate) exercise intensity by matching their efforts to particular RPE levels (such as 9, 13, and 17). The results show that older children and adolescents perceive their levels of exertion as accurately as adults (Eston & Williams, 1986; Miyashita *et al*, 1986). Recently, both Ward *et al* (1991) and Williams *et al* (1991) have suggested independently that young children (aged 8-14 years) were also capable of interpreting and using the RPE scale to control (or *produce*) exercise levels. In contrast, however, Bar-Or (1977) reported that children below age 16 were less competent using the RPE than subjects older than 18. Further, it has been observed that children's success (relative to that of adults) with the RPE seems to be age-related and that there is possibly a critical threshold at 9 years, below which children are unable to use the scale effectively (Miyashita *et al*, 1986).

Such inconsistencies have prompted the suggestion that difficulties in comprehension with the existing RPE scale may be influencing findings and that a scale with a narrower range and more 'appropriate' expressions of effort would be more meaningful to younger subjects (Williams *et al*, 1991). Accordingly, the Children's Effort Rating Table (CERT) has been devised (see Figure 3). The purpose of the present study was to collect preliminary information on the validity and reliability of the CERT as a means of controlling exercise intensity during cycle ergometry among young children.

Method

Subjects

Sixteen children, aged 8-11 years, were selected from a primary school in Liverpool, England. Descriptive characteristics of the children are presented in Table 3. Prior to any testing, parents or guardians had provided written informed consent for their children's involvement in a series of cycle ergometer exercise bouts to be conducted on the school's premises.

Table 3.
Sample Characteristics

Measure	Boys, n = 8		Girls, n = 8		All, N = 16	
	M	*SD*	*M*	*SD*	*M*	*SD*
Age (years)	9.9	1.2	10.0	1.0	10.0	1.1
Stature (m)	1.37	0.07	1.36	0.08	1.36	0.08
Mass (kg)	31.0	5.4	32.1	6.8	31.6	6.0
Resting heart rate (b/min)	89.8	5.0	96.1	11.2	92.9	9.2

The Children's Effort Rating Table (CERT).

The CERT scale (Williams *et al*, 1994) evolved with developmentally appropriate verbal language and a numerical scale thought to represent better (than that used with adults) the types of heart rates commonly found among children. The numerical range of the CERT (1 to 10) reflects a conceptual model[1] in which perceived effort and heart rate (in the range 100 to 200 b/min) are linearly related as follows:

$$\text{Heart Rate} = 100 + 10x$$

where x is the CERT value reported at any one time.

Procedure

One week prior to testing, the children were introduced to the CERT scale and the exercise equipment. Following an explanation of how the verbal

[1] The model stated above was actuallly inaccurately communicated to me in 1993. The 'correct' model depicted in Williams *et al* (1994) yields the following equation: Heart Rate = 110 + 12(CERT). This is expanded upon on p. 75 of this thesis.

expressions should be interpreted in numerical form, each child was given a copy of the CERT to study and keep. Data collection subsequently took place on three occasions, one week apart. In *Stage I*, subjects performed a continuous incremental exercise test on a mechanically braked cycle ergometer (Monark 814), commencing at 25 W for four minutes (at 'steady state') and increasing thereafter by 25 W (or 10 W for the youngest children) after every four minutes. Heart rate was monitored by telemetric means (Polar PE3000 Sport Tester), and recorded in the final 30 s of each bout, as was the rating of effort. The experimenter stopped the exercise test as soon as each child reached a CERT value of 9 or 10.

Stage II required subjects to adjust their power output (exercise intensity) to three specific CERT levels; 5, 7, and 9 (randomly ordered for each child). Following a 4-min, 25 W warm-up, the experimenter was instructed by the child to vary the power output as appropriate in order that he/she could achieve and maintain each CERT level for four minutes. The objective indicators of effort, heart rate and power output, were recorded during the final 30 s at each CERT level. *Stage III* involved a repeat of Stage 2, to estimate test-retest reliability.

Data Analysis

Regression analysis was employed to assess the criterion validity of the CERT, that is, how well CERT ratings could predict heart rate and power output. Pearson correlation coefficients were calculated to assess the ability of subjects to produce specific (objective) exercise intensities based on their perceptions of effort. The reliability of these findings was examined via the intraclass correlation technique (see Thomas & Nelson, 1990).

Results

Stage I

All the children provided CERT responses for at least three different work loads, most of whom (69%) reached number 10 on the scale. Table 4 highlights statistically significant ($p < .01$) positive associations between perceived (CERT) and actual effort (heart rate and power output). Linear regression analyses yielded the predictive ability of CERT as follows:

Heart Rate = 127.7 + 6.6 (CERT); SE = 18.3 b/min

Power Output = 26.7 + 6.8 (CERT); SE = 19.7 watts

Regression techniques were then employed to estimate individual heart rate and power output values for CERT levels 5, 7, and 9 for use in Stage II.

Table 4.
Correlations* Between Perceived and Objective Effort

Variable	Boys	Girls	All
CERT and Heart Rate	0.70	0.83	0.76
CERT and Power Output	0.70	0.82	0.75

*$p < .01$

Stage II

Power outputs and heart rates produced by the children at CERT levels 5, 7, and 9 correlated well with those estimated for the corresponding levels in Stage I: for power output, *rs* = 0.84, 0.87, and 0.91; for heart rate, *rs* = 0.65, 0.78, and 0.79 for all subjects ($p < .01$). However, analysis of variance indicated that both mean power output and heart rate values were significantly *lower* ($F_{1,45} = 59.4$ and $F_{1,45} = 49.5$; $p < .01$) in Stage II than those predicted from Stage I.

Stage III

The relationships observed in Stage II were generally reliable with exercise intensities produced at each CERT level correlating highly with those in Stage III (Table 5). The *overall* reliability coefficients were $R = 0.96$ and 0.78 for power output and heart rates, respectively. Mean values for power output were not significantly different ($F_{1,14} = 0.0$; $p > .05$) across the two trials, as were those for heart rate ($F_{1,14} = 1.0$; $p > .05$; see Appendix A1).

Table 5.
Test-Retest Intraclass Correlations* Between Stage II and Stage III Exercise Intensities

VARIABLE	Boys CERT LEVEL 5	7	9	Girls CERT LEVEL 5	7	9	All CERT LEVEL 5	7	9
Power Output	.97	.98	.90	.90	.93	.90	.95	.97	.91
Heart Rate	.72	.71	.67	.96	.78	.82	.86	.65	.77

*$p < .05$.

Discussion

Data from the present study suggest that the recent perceived effort scale designed for use with children (CERT) is a valid indicator of physiological effort during cycle ergometry. Children aged 8-11 years were not only successful in perceiving experimenter-applied changes in their exercise efforts, but could also reliably regulate such effort (in line with criterion values) by applying their understanding of the CERT. Further, whilst this ability was prone to some underestimation (the objective effort produced at CERT levels 5, 7, and 9 was about 13 to 18% lower than expected), it was repeatable and apparently grasped both by boys and girls.

Previously, Ward & Bar-Or (1990) employed Borg's RPE 6-20 scale in a similar design to that of the present study and found that their sample of overweight children (aged 9-15 years) produced overestimated efforts at low intensities (RPEs 7 and 10) and underestimated efforts at the highest intensity (RPE 16). An explanation for these 'errors' of perception, which have also been observed in studies of adults (see Van den Burg & Ceci, 1986) has been offered by Noble (1982), who argued that the psycho-physical process of reproducing a given exercise effort from memory is not the same as that of estimating effort intensity during ongoing exercise. Likewise, the psycho-physical process of producing effort in an incremental manner may be different to that of producing effort in a decremental, or even intermittent manner; those subjects required (by random selection) to produce levels of effort increasing from low to high (i.e., CERT 3-5-7-9 or RPE 7-10-13-16) may be more successful than other subjects required to produce levels of effort decreasing from high to low (CERT 9-7-5-3 or RPE 16-13-10-7), and less successful than those allowed recovery periods between exercise levels. This methodological consideration warrants future examination.

In line with the conceptual model of the CERT, high heart rates, for example 170, 190, and 200 b/min, were well predicted by CERT values of 7, 9, and 10 (174, 187, and 194 b/min, respectively). However, the heart rate : CERT relationships was less accurate at moderate or light intensities; CERT values 2, 3, and 5 respectively predicting heart rates of 140, 148, and 161 b/min. In addition, the possibility that the model (and its applications) may be affected by gender differences is raised in this study. Further research is required to elucidate these preliminary results.

The first data involving the CERT have recently emerged from an exploratory study with children aged 4 to 8 years (Williams *et al,* 1993). Whilst the criterion validity of the CERT was confirmed (even, to some extent, among the youngest children), this group were generally unable to regulate their efforts accurately during an incremental stepping task. It was suggested that this may have been caused by the children's limited experience of working at varied exercise intensities. With the older children of the present study, it is reasonable to assume that they have experienced a broader range of intensities, although still not sufficient to be unerring in their perception of effort. A consequence of this lack of 'training' is often an inability to sustain exercise beyond brief, over-vigorous bouts, although this may also be common among adults returning to exercise after prolonged periods of abstention.

This study has reinforced the potential of the CERT as a tool capable of quantifying young children's sense of effort, and regulating their exercise output. Owing to the small sample size used, however, the present findings pertinent to both the estimation and production trials merit verification by a larger study. Additionally, since it was designed to replace the RPE scale, the credibility of the CERT needs to be addressed via a comparative investigation. The following study will deal with these concerns.

STUDY 2

Children's Ratings of Effort During Cycle Ergometry: An Examination of the Validity of Two Effort Rating Scales[3]

[3] The content of this chapter appeared in the paper by Lamb, K.L. (1995), *Pediatric Exercise Science*, 7, 407-421.

Introduction

Research into the usefulness of perceived effort rating scales in exercise evaluation and prescription has been abundant over the past twenty years. The 6-20 Rating of Perceived Exertion (RPE) psychophysical scale established by Borg (1970; 1985) has been so dominant during this time that exercise scientists have seldom considered using an alternative for their numerous investigations into the perceived effort (exertion) concept. Thorough reviews have been published which establish and reinforce the validity and applications of this single item amongst adults (Borg & Noble, 1974; Pandolf, 1983; Watt & Grove, 1993).

More recent research has focused on the ability of *children* to perceive accurately changing levels of exercise intensity. The relationship between perceived and objective (heart rate and/or power output) levels of intensity has been found to vary in magnitude; r = 0.64 to 0.94 among children above 9 years of age (Gillach *et* al, 1989), and r = 0.55 to 0.74 for children aged 7-9 years (Bar-Or, 1977; Miyashita *et al*, 1986). Gillach *et al.* (1989) argued that this disparity in validity correlations obtained with Borg's RPE scale (Figure 2, p. 24) is a consequence of the type of statistical method employed and that most reports have probably *over-estimated* the degree of relationship between perceived effort and physiological strain. Other authors (Bar-Or, 1977; Bar-Or & Ward, 1989) have highlighted how children tend to under-rate exercise intensity compared to adults, and thereby do not exhibit the kind of perceived effort/heart rate ratio (1:10) synonymous with the concept of the RPE scale. Such uncertainty concerning RPE's applicability to the exercising child provides a case for a child-specific effort rating scale, which has an alternative notation (Bar-Or & Ward, 1989) and/or a narrower range of numbered responses and more meaningful expressions of effort than the RPE scale (Ward *et* al, 1991; Williams *et al*, 1991).

A development of this kind has recently taken place (see Figure 3, p.24) in the form of the Children's Effort Rating Table (CERT) devised by Williams *et al.* (1994). It can be seen that compared to Borg's scale, CERT has five fewer possible responses, a range of numbers (1-10) more familiar to children (than 6-20), and verbal expressions that not only accompany all 10 numbers, but represent words chosen and understood by children as descriptors of exercise effort (Williams *et al,* 1994). Support for the validity of this new scale has emerged from two pilot studies conducted among schoolchildren aged 5 to 9 years (Williams *et al*, 1993) and 8 to 11 years (Study 1). These studies reported validity correlations (against heart rate) of $r = 0.76$ during incremental cycle ergometry (Study 1) and $r = 0.73$ to 0.99 during incremental stepping (Williams *et* al, 1993). As a logical advancement, therefore, the primary purpose of the present study is to reassess and compare the validity, and test-retest reliability, of children's effort ratings using the RPE and CERT scales. It will also emulate the approach of Gillach *et al.* (1989) and address how the manner of primary data analysis can affect the way in which the results are interpreted.

Method

Subjects

Data were collected on seventy children (28 boys and 42 girls) representing 96% of Year 5 (4th Grade) of a Primary school in Chester, England. Initially, 73 children had been randomly assigned to one of two groups: Group 1 (RPE) and Group 2 (CERT)[1]. Three subjects were subsequently unavailable owing to absence from school. Prior to any testing, written informed consent was obtained from the parents or guardians of each child (see Appendix B1).

Procedure

Three days before testing the children were introduced by a teacher to both effort rating scales (RPE and CERT) and the exercise equipment. Following an explanation of how the verbal expressions for each scale should be interpreted in numerical form, the children in each group were given copies of their designated scale to study and keep. Data collection for both groups subsequently took place on the school premises on two occasions, one week apart. On the first occasion, Trial 1 (T1), each group performed a continuous incremental exercise test on a mechanically braked cycle ergometer (Monark 814), commencing at 25

---

[1]A random two-group design was adopted in preference to a single-group (cross-over) design owing to the author's concern for the potential negative aspects of repeated measures, such as *latency*, *carry-over*, and *fatigue* effects (Munro & Page, 1993; pp. 171-172). It is acknowledged that had the sample sizes been small, the present design would not have been appropriate.

Table 6.
Subject Characteristics (Means and Standard Deviations).

	GROUP											
	CERT (*n* = 36)						RPE (*n* = 34)					
Variable	Overall		Boys		Girls		Overall		Boys		Girls	
	M	*SD*	M	*SD*	M	*SD*	M	*SD*	M	*SD*	M	*SD*
Age (yr)	9.6	0.3	9.6	0.3	9.6	0.3	9.6	0.3	9.6	0.3	9.6	0.4
Height (m)	1.35	0.05	1.35	0.04	1.35	0.05	1.36	0.05	1.36	0.05	1.35	0.05
Mass (kg)	31.39	4.92	30.56	2.42	32.05	6.24	31.94	4.80	31.17	2.55	32.36	5.68
[1]Peak power (watts)	99.9	17.5	109.2	11.7	92.8	18.1	104.9	15.0	111.5	15.9	101.2	13.4

[1] At a predicted maximum heart rate (220 - age), estimated by regression analysis of Trial 2 data.

W and increasing thereafter by 25 W. All stages lasted for 4 min. Pedal rate was held constant at 50 rpm. Prior to exercise, each subject was re-introduced to his/her assigned effort rating scale and informed in a standardised manner (see Appendix B2) what the scale was measuring and how it was to be used (by pointing to the appropriate number or verbal expression when requested by the experimenter) during the exercise trial. Heart rates were monitored using radio telemetry (Sports Tester PE4000, Polar Electronics, Finland) and recorded during the final 15 s of each exercise stage, as were the ratings of perceived effort. Testing was terminated as soon as each child reported either an RPE of 18 or a CERT of 9, or above, or could not maintain the required power output for more than one minute. Trial 2 (T2) was a repeat of T1 and was performed to assess the test-retest reliability of each scale.

Data Analysis

Data were analyzed using the MANOVA procedure in SPSS PC+ (V.4) statistical software (SPSS, 1990) to reveal differences in heart rate and perceived effort variability. With this model the *within-subject* (repeated measures) factors (exercise levels and trials) are treated as dependent variables, and the *between-subject* factors (group and sex) are treated as independent variables. This multivariate approach is preferable since it diminishes the problem associated with univariate repeated measures designs that lack the important pre-requisite of *sphericity* (Schutz & Gessaroli, 1987), namely the increased probability of making a Type I error. In addition, to reduce the risk of inflating the experimenterwise error associated with analysing more than one dependent variable belonging to the same data set (heart rate and perceived effort), the α-level was set at .025. Post-hoc comparisons of specific factor levels were analysed, where appropriate, with the Tukey test.

Interclass correlations (Pearson's r), used to quantify the predictive validity of each scale, were derived from two methods; **Method I** involved determining the coefficients between the perceived and objective measures of exercise intensity for every individual subject, then within both groups the mean of these was calculated. **Method II** involved the simultaneous analysis of all subjects' data (in each group) to give a single, overall correlation coefficient. Unlike Method I, this approach gives a comprehensive representation of how two variables interrelate for a sample as a whole. A two-tailed Fisher z transformation (Morehouse and Stull, 1975, pp. 203-204; see Appendix D2) was employed with each method to test whether correlation coefficients differed significantly ($p <$ .05) between groups and across trials.

Intraclass correlations (Safrit & Wood, 1989, pp. 52-59) obtained from a two-way analysis of variance model (in which trial-to-trial variance was accounted for) were calculated to give a quantitative indicator of the overall test-retest reliability (consistency) of each scale (for example, see Appendix A2).

Results

Univariate analysis of variance revealed that members of the CERT group were not significantly different ($p > .05$) in terms of their biometric data (see Table 6 and Appendix B3) from members of the RPE group. However, sex main effects were observed for estimated peak power, $F(1, 63) = 13.9$, $p < .01$, with the boys having a higher mean peak power. Exercise heart rates over the two trials ranged from 106-218 bpm (50-103% of age-predicted maximum) for the CERT group and 107-223 bpm (51-106% of age-predicted maximum) for the RPE group, with all subjects reaching a peak rate in excess of 162 bpm (77%) in each trial. Sixty-seven subjects completed T1 and T2 up to an intensity of 75 W, but only 22 reached the fourth level (100 W) on both occasions. Accordingly, unless otherwise indicated, the following analyses are limited to data collected at three exercise levels (25, 50 and 75 W).

The physiological (heart rates) and psychological (perceived effort ratings) responses to the standard workloads over two trials are presented in Tables 7 and 8, respectively. For *heart rate* data, the scale by trials interaction was found to be significant, $F(1, 63) = 8.5$, $p < .01$ (Appendix B4). Post-hoc analysis indicated that the RPE group had significantly lower mean heart rates ($T = 6.1$, $p < .05$) in T2 than in T1 (see Figure 5). Significant main effects were found for levels of exercise, $F(2, 126) = 1383.0$, trials, $F(1, 63) = 13.1$, and sex, $F(1, 63) = 14.4$.

For *perceived effort* data, a significant 2-factor interaction effect was found for trials by exercise levels, $F(2, 126) = 16.0$, $p < .01$ (Appendix B4).

Table 7.

Mean (SD) Heart Rate Responses to Four Standardised Exercise Levels Over Two Trials (T1 & T2)

Exercise Level	ALL		BOYS		GIRLS	
	T1	T2	T1	T2	T1	T2
25 WATTS						
CERT	131.1 (15.9)	132.4 (14.5)	123.1 (10.4)	125.1 (12.1)	137.0 (16.9)	137.9 (14.0)
RPE	134.6 (13.1)	128.7* (13.6)	131.3 (11.5)	123.0* (11.2)	136.5 (13.9)	132.0 (14.0)
50 WATTS						
CERT	160.5 (17.6)	159.1 (17.6)	150.4 (12.2)	148.0 (10.1)	168.2 (17.4)	167.4 (17.6)
RPE	163.8 (15.4)	155.8* (14.3)	158.3 (15.5)	150.6* (11.8)	167.0 (14.8)	158.7* (15.0)
75 WATTS						
CERT	187.3 (13.6)	184.6* (13.3)	180.5 (10.7)	178.1 (7.9)	192.6 (12.9)	189.8 (14.5)
RPE	187.5 (13.2)	182.6* (12.7)	181.8 (15.1)	177.3* (12.8)	190.9 (11.0)	185.6* (11.9)
100 WATTS						
CERT	196.4 (7.4)	195.4 (7.2)	194.4 (7.4)	195.4 (7.2)	- -	- -
RPE	198.2 (12.2)	196.3 (9.9)	192.5 (11.9)	190.3 (6.4)	203.1 (11.0)	201.4 (9.7)
OVERALL						
CERT	162.2 (28.4)	161.4 (27.1)	158.5 (28.9)	157.9 (27.9)	165.5 (27.2)	164.6 (26.2)
RPE	166.2 (27.2)	160.4* (27.8)	162.2 (26.6)	156.0* (27.4)	168.6 (27.4)	163.0* (27.9)

* significantly different ($p < 0.05$) to corresponding T1 value.

Table 8.

Mean (SD) Perceived Effort Responses to Four Standardised Exercise Levels Over Two Trials (T1 & T2).

Exercise Level	ALL		BOYS		GIRLS	
	T1	T2	T1	T2	T1	T2
25 WATTS						
CERT	2.62 (1.03)	1.97* (0.78)	2.53 (0.74)	2.00 (0.84)	2.70 (1.22)	1.95* (0.76)
RPE	9.39 (2.19)	7.97* (2.02)	8.50 (1.93)	7.00* (1.35)	9.90 (2.21)	8.52* (2.16)
50 WATTS						
CERT	4.17 (1.27)	3.66* (1.26)	4.07 (1.28)	3.33* (1.11)	4.25 (1.29)	3.90 (1.33)
RPE	11.36 (2.32)	10.24* (2.42)	10.92 (2.11)	9.33* (2.15)	11.62 (2.43)	10.76* (2.47)
75 WATTS						
CERT	5.91 (1.56)	6.11 (1.98)	5.67 (1.59)	5.40 (1.68)	6.11 (1.56)	6.68 (2.06)
RPE	13.33 (2.89)	13.45 (3.08)	12.92 (3.40)	13.25 (2.93)	13.57 (2.62)	13.57 (3.23)
100 WATTS						
CERT	8.22 (1.86)	8.22 (1.86)	8.22 (1.86)	8.22 (1.86)	- -	- -
RPE	14.31 (3.66)	14.00 (3.03)	15.17 (3.71)	14.83 (1.60)	13.57 (3.74)	13.29 (3.86)
OVERALL						
CERT	4.53 (2.15)	4.23* (2.48)	4.78 (2.36)	4.35* (2.55)	4.32 (1.93)	4.14 (2.42)
RPE	11.71 (3.16)	10.96* (3.51)	11.40 (3.50)	10.57* (3.64)	11.88 (2.95)	11.19* (3.44)

* significantly different ($p < 0.05$) to corresponding T1 value

Post-hoc analysis indicated that effort ratings in T2 were significantly lower (T = 0.6, $p < .01$) than in T1 at the 25 and 50 W intensity levels (see Figure 6). Main effects for scale, $F(1, 63) = 295.2$, exercise levels, $F(2, 126) = 219.3$, and trials, $F(1, 63) = 17.0$, were also significant.

Validity

Table 9 shows both effort rating scales to have strong positive associations with objective measures of exercise intensity when data were analyzed via Method I. Only among girls in T1 were correlations significantly ($p < .05$) larger for the group using the CERT scale than those using the RPE. Pearson coefficients derived via Method II (Table 10) were markedly and consistently lower than those in Method I, and in T1, RPE coefficients were now significantly lower than CERT. The predictive validity of each scale (in T2) is represented mathematically as follows:

CERT

Heart Rate = 127.63 + 8.1 (CERT); $r^2 = 0.53$, SEE = 19.1 bpm

Power Output = 20.8 + 7.9 (CERT); $r^2 = 0.64$, SEE = 14.7 watts

(Appendix B5)

Figure 5.
Interaction of Group by Trials on Heart Rate

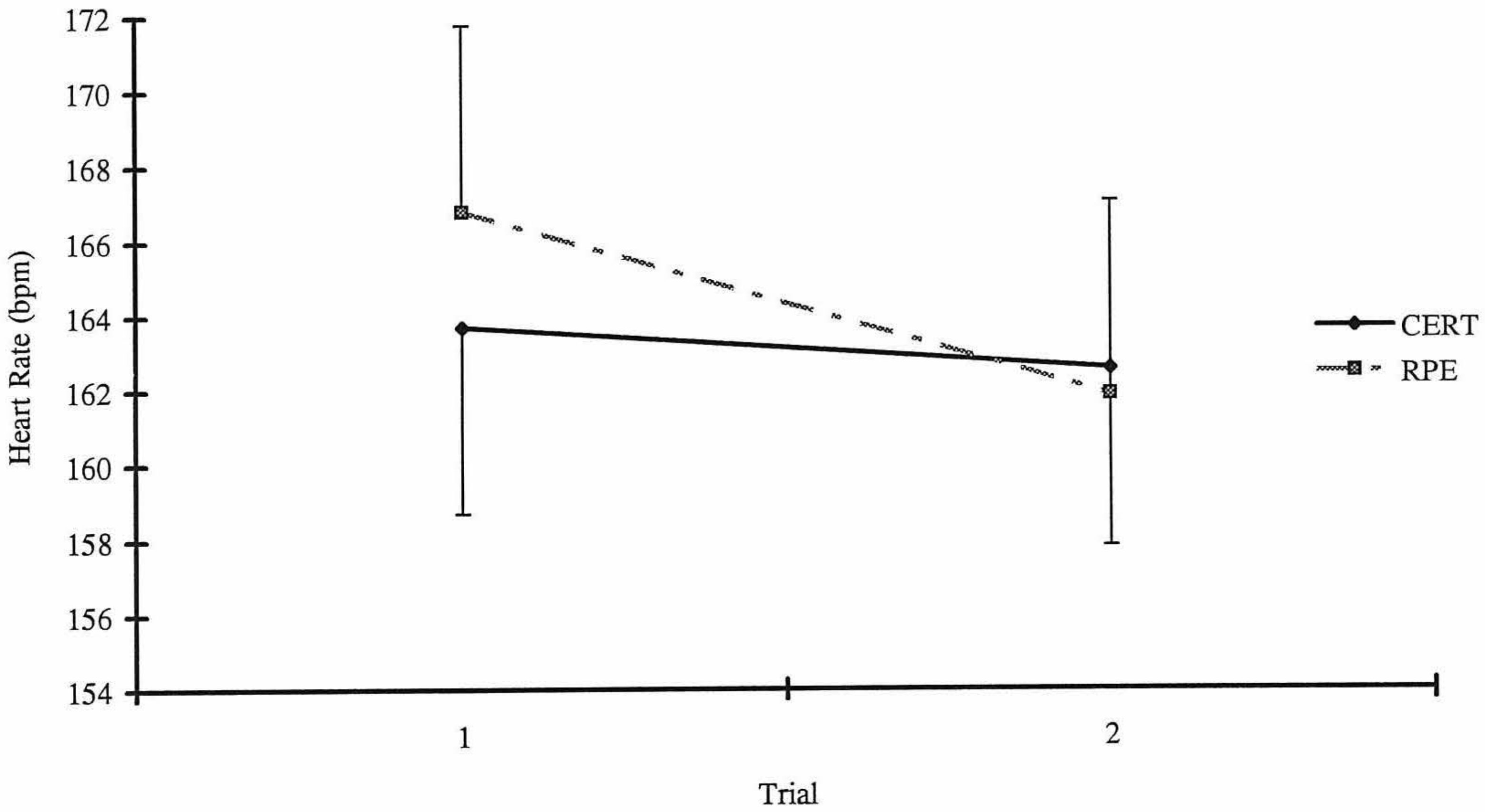


Figure 6.
Interaction of Trials by Exercise Levels on Effort Ratings

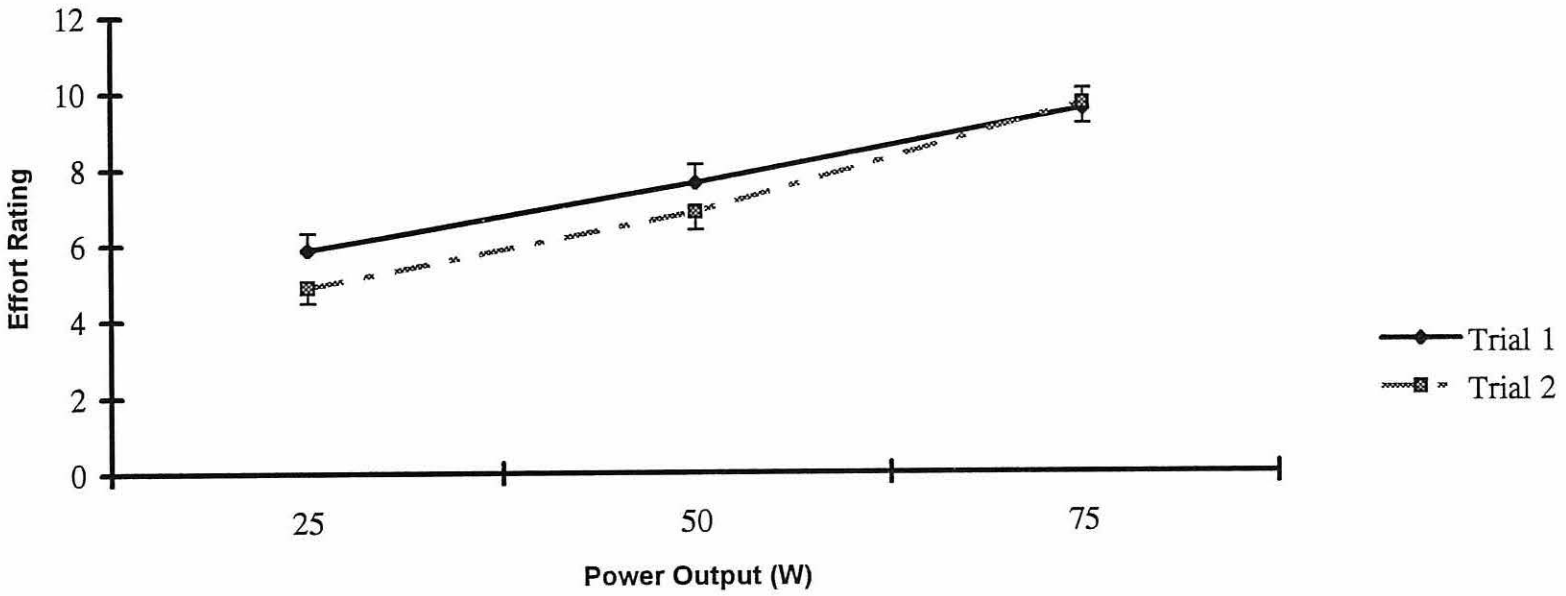


Table 9.
Pearson Correlations[1] Between Perceived Effort Ratings and Objective Measures of Exercise Intensity: Method I

RELATIONSHIP	EFFORT RATING SCALE			
	RPE		CERT	
	T1	T2	T1	T2
ALL				
Perceived Effort and Heart Rate	0.90	0.95	0.96	0.96
Perceived Effort and Power Output	0.93	0.97	0.98	0.97
BOYS				
Perceived Effort and Heart Rate	0.93	0.95	0.94	0.96
Perceived Effort and Power Output	0.96	0.97	0.97	0.97
GIRLS				
Perceived Effort and Heart Rate	0.88	0.94	0.97*	0.96
Perceived Effort and Power Output	0.92	0.96	0.98*	0.97

[1] All correlations significant ($p < 0.001$)
* Significantly different ($p < 0.05$) to corresponding RPE coefficient.

Table 10.
Pearson Correlations[1] Between Perceived Effort Ratings and Objective Measures of Exercise Intensity: Method II

RELATIONSHIP	EFFORT RATING SCALE			
	RPE		CERT	
	T1	T2	T1	T2
ALL				
Perceived Effort and Heart Rate	0.50	0.64	0.69*	0.73
Perceived Effort and Power Output	0.59	0.70	0.80*	0.80
BOYS				
Perceived Effort and Heart Rate	0.58	0.79	0.79*	0.79
Perceived Effort and Power Output	0.68	0.83	0.84	0.84
GIRLS				
Perceived Effort and Heart Rate	0.45	0.54	0.64	0.69
Perceived Effort and Power Output	0.53	0.63	0.75*	0.78

[1] All correlations significant ($p < 0.001$)
* Significantly different ($p < 0.05$) to corresponding RPE coefficient.

RPE

Heart Rate = 107.1 + 4.9 (RPE); r^2 = 0.40, SEE = 21.8 bpm

Power Output = 1.5 + 5.0 (RPE): r^2 = 0.49, SEE = 18.6 watts

Boys' effort ratings in each group were found to be better predictors of heart rates (CERT: r^2 = 0.61, SEE = 17.6; RPE: r^2 = 0.61, SEE = 17.4) than girls' (CERT: r^2 = 0.46, SEE = 19.7; RPE: r^2 = 0.28, SEE = 23.9). With regard to power output, sex differences were only noticeable in the RPE group (Boys: (r^2 = 0.69, SEE = 15.1; Girls: r^2 = 0.39, SEE = 20.2).

Reliability

Although mean perceived effort ratings in both groups were found to differ significantly ($p < .05$) between T1 and T2, in neither group did this amount to more than a fraction of one unit. Table 11 shows both scales to have acceptable test-retest reliability (in the context of this study), though clearly this is not consistent for every exercise level. This pattern was maintained within each sex group.

Discussion

This study, the first to compare two effort rating scales, has confirmed that using *either* scale children's perceptions of effort do represent valid reflections of actual (and changing) physiological effort during continuous,

Table 11.
Test-Retest Reliability Analysis (Intraclass *R*) of Perceived Effort Ratings

Exercise Level	EFFORT RATING SCALE	
	RPE	CERT
Overall		
Boys	0.90	0.93
Girls	0.89	0.89
All	0.90	0.91
25 Watts		
Boys	0.43	0.23
Girls	0.73	0.12
All	0.70	0.14
50 Watts		
Boys	0.75	0.50
Girls	0.82	0.69
All	0.81	0.62
75 Watts		
Boys	0.88	0.86
Girls	0.86	0.82
All	0.86	0.84
100 Watts[1]		
Boys	0.83	0.83
Girls	0.95	-
All	0.90	0.83

[1] Only 13 children in the RPE group and 9 children in the CERT group (all boys) completed this level twice.

incremental cycle ergometry. However, it should be stated clearly that for some individuals, perceptions of effort seem to be poorly or inconsistently related to changing physiological demands. As an example, one child using the CERT scale (in T1) rated her efforts at each workload as follows (heart rates in brackets): 25 W, 2 (123); 50 W, 3 (151); 75 W, 3 (188); 100 W, 5 (206). Using the RPE scale another child rated her efforts as follows: 25 W, 8 (132); 50 W, 15 (168); 75 W, 11 (188); 100 W, 14 (195).

Before expanding on the validity of the two scales, their reliability must be considered. A reliable scale should yield consistent effort ratings on a test-retest basis given that the testing procedures are held constant. Ward & Bar-Or (1987) have reported a two-day test-retest correlation of 0.86 for the RPE scale with a group of 20 obese children (aged 9-15 years), and elsewhere on the same sample (1989) correlations (presumably Pearson product-moment) ranging from 0.59 to 0.92 for specific exercise intensities, with the higher coefficients pertaining to the higher exercise intensities. The more appropriate statistic for univariate test-retest reliability, is the intraclass coefficient, obtained from a repeated measures analysis of variance (Safrit & Wood, 1989) and calculated as R = (MSs - MSw)/ MSs, where MSs represents *between-subject* variance, and MSw represents *within-subject*[2] variance. No previous studies involving children have stipulated this statistic as their measure of reliability.

---

[2]Where MSw = $(SS_{Trials} + SS_{Interaction}) / (df_{Trials} + df_{Interaction})$.

In the present study, the overall intraclass coefficients for both scales were high (≥0.90), even though the mean rating values differed significantly between tests. These findings reflect that significant trial-to-trial fluctuations in subjects' mean ratings occurred, whereas the *interaction* variance (reflecting the pattern or order of ratings across trials) was relatively low. This latter finding was also observed at each exercise intensity level. However, the occurrence of limited between-subject variance among children using the CERT scale at the two lowest intensities (25 and 50 W), and to a lesser extent those using the RPE scale, meant that the impact of the accompanying within-subject variance was inflated, causing the intraclass coefficient to be markedly reduced (see Table 11). As exercise intensity increased, between-subject responses became more variable (as exercise tolerances became challenged), whilst within-subject variance remained relatively low. Accordingly, intraclass *R*s increased.

Post-hoc analysis of the scale by trials interaction effect on *heart rate* revealed a significant decrease between trials in the RPE group only, and is thereby difficult to explain. The significant main effect of sex on heart rates (heart rates of girls being higher than boys at equivalent workloads) might be due to natural, sex-related differences in the autonomic nervous regulation of the heart (Bar-Or, 1983), or possibly a consequence of the boys' superior estimated exercise capacities (peak power). If the latter is true, since the girls did not have higher perceived effort ratings to match their higher heart rates (and greater relative effort), they have seemingly underestimated the physiological stress imposed upon them.

The significant interaction effect of trials by exercise levels on *perceived effort* suggests that perceptual interpretations of low levels of exercise intensity (in this case 25 and 50 W) may have been due to a protocol-related practice (or testing) effect in which the children adjusted their perceptual responses for T2 on the basis of their experiences of T1. A third identical trial might have elucidated this finding.

Despite the consistency of ratings at the higher levels of exercise, there existed an apparent reluctance among the RPE users to rate effort in excess of 14 or 15, even though their corresponding heart rates averaged approximately 95% of age-predicted maxima. This apparent *under-estimation* confirms the earlier work of Bar-Or (1977) who for exercise at almost the same relative exercise intensity, reported average peak RPEs among 189 10-12 year-old boys of approximately 15.5. A possible cause of these findings is that children so young are unable to fully comprehend the RPE scale's verbal expressions and/or number series. A recent unpublished study (Tones, 1995) has revealed that 78% of a class of 28 9- to 10-year-old boys and girls could not comprehend the phrases "NO EXERTION AT ALL" and "MAXIMAL EXERTION" as they appear in the RPE scale. In contrast, all of the children were able to interpret the words used in the CERT scale.

Clearer differences between the two scales are apparent when their overall predictive validities are addressed. Nonetheless, had the alternative method of correlational analysis (described as Method I) been adopted, claims for a spuriously inflated level of validity for both scales, and in particular the RPE, could now be made. When data were analyzed using Method II, the validity correlations for both scales decreased, but those for CERT were now distinctly superior to those for RPE. These changes due to the type of statistical analysis substantiate the thesis of Gillach *et al.* (1989), who declared (albeit without specific references) that many previous investigators were guilty of not stating their methods of statistical analyses and inclined to adopt the method of calculating mean values from individual coefficients (Method I). In their own study using the RPE scale with 283 children (aged 10 to 14 years) and 295 adults during cycle ergometry, Gillach *et al.* did indeed show reductions in HR-RPE correlations from 0.94 to 0.65 (children) and 0.95 to 0.64 (adults) due to the adoption of Method II. Undoubtedly there is a case for researchers in this field to be more explicit when reporting their statistical procedures, as exemplified by Arnhold, Ng & Pechar (1992).

The smaller, but appropriate, correlations obtained for the CERT scale are not dissimilar in magnitude to those reported recently for a smaller sample of 8-11 year old school children during cycle ergometry (Study 1). In that study CERT correlated 0.76 and 0.75 with heart rate and power output, respectively. Likewise, RPE correlations in the present study, especially in T1, are comparable to those previously reported for 7-9 year-olds (Bar-Or, 1977; Miyashita *et al*, 1986). The finding that the relationships between CERT and indices of objective effort were more stable between T1 and T2 (regardless of sex) than those for RPE, may indicate that the CERT scale is preferable where practice is not available. Other reports showing the effect of relatively short-term practice on the RPE-objective effort relationship during these so-called *estimation* trials (in children *or* adults) are scarce. One study (Gillach *et al*, 1989) has reported almost identical HR-RPE correlations for two repeated trials (0.64 and 0.65 for children, 0.63 and 0.64 for adults), but since they were one year apart it is difficult to accept that the first trial served as a practice for the second.

Despite the comparative success of CERT as an indicator of physiological exertion, 48% of the variance in heart rates remains unexplained by its ratings of effort. Consequently, the present data do not fit too well the conceptual model of the CERT scale (Williams *et* al, 1994, see Appendix B6), which (approximately) relates heart rate and perceived effort linearly in the form: Heart Rate = 110 + 12(CERT). In theory, CERT values of 3, 5, 7, and 9 should equate to heart rates of 146, 170, 194 and 218 bpm, respectively. Data from T2 fit a linear model (Appendix B7) which predicts heart rate values of 151, 168, 184, and 201 bpm, revealing a larger disparity with the conceptual model at the higher exercise intensities (a slightly better fit was observed for boys' data, whilst the converse was true for girls'). Accordingly, a more appropriate description of the present data is actually a *curvilinear* model (see Appendix B8): HR = [120.6 +

6.8(CERT) + 1.75($CERT^2$) - 0.18($CERT^3$)], which was observed to yield a superior r^2 of 59% (and 69% for a similar model for boys). These deviations from theory corroborate the findings from the previous study (Study 1) which predicted heart rates of 146, 161, 174, and 187 bpm for the same four CERT levels.

In terms of work (power) output, the model of Williams *et al.* (1994) predicts values of 40, 60, 80, and 100 W at CERT levels 3, 5, 7, and 9, respectively. This is based on the relationship: Power Output ≈ 10 + 10(CERT). The present data (from T2) fit a linear model which predicts power outputs of 45, 60, 76, and 92 W for the same CERT values (r^2 = 64%). Again, these values follow the same pattern as those predicted in the aforementioned pilot study (47, 61, 74, and 88 W), and reinforce the non-linear trend exhibited by the heart rate data.

The suitability of the RPE conceptual model (Borg, 1970), Heart Rate = 10(RPE), to children during progressive exercise has not been adequately addressed. RPE ratings in the present data explain only 40% of heart rate variance in a linear model that predicts heart rates of 146, 166, 181, and 195 bpm for RPEs 8, 12, 15, and 18, respectively. On this evidence, particularly at low intensities, Borg's conceptual model is clearly inappropriate for , exercising children. Furthermore, Bar-Or & Ward (1989), reporting on data first published in 1977 (Bar-Or, 1977), have revealed that the 1:10 RPE-Heart Rate ratio did not apply with 10- and 13-year old boys cycling at intensities ranging from 50-175

W. More precisely, the RPE-Heart Rate ratios were similar to those observed in the present study, that is, (for boys) 0.71:10 and 0.69:10 in T1 and T2, respectively, and increasing from 0.65:10 at 25 W to 0.79:10 at 100 W (T1), and from 0.57:10 at 25 W to 0.81 (T2). Ratios for girls are very similar, though they do span a narrower range across the four workloads. These findings support the argument for a child-specific model to assess perceived exertion (Ward *et al*, 1991).

The importance of the present study emerges when one contemplates the potential for using effort perceptions for *regulating* or *prescribing* children's exercise levels, particularly in a physical education setting (Eston, 1984; Williams *et al*, 1991). Previous laboratory-based attempts to get children to execute accurately RPE-prescribed exercise intensities, either by self-adjustment (Ward and Bar-Or, 1990; Williams *et al*, 1991) or experimenter-adjustment (Ward *et al*, 1991) have achieved some degree of success. Similarly, the findings from Study 1 suggested that children can use the CERT scale to regulate exercise effort. Further research into the validity of employing both scales in this production mode is presented in the following study.

STUDY 3

Exercise Regulation During Cycle Ergometry Using the CERT and RPE Scales[4]

[4] The content of this chapter appeared in the paper by Lamb, K.L. (1996), *Pediatric Exercise Science*, 8, 337-350.

Introduction

Investigations into the ability of children to use effort rating scales to self-regulate accurately, or *produce* pre-determined levels of exercise intensity, have begun to emerge recently in the exercise science literature (Ward & Bar-Or, 1990; Ward *et al*, 1991; Williams *et al*, 1991; Ward *et al*, 1995). The rationale for such research was founded on the recognition that left to their own devices, children readily misjudge the intensity dimension of exercise and consequently do not adhere to levels suitable for yielding optimal cardio-respiratory and associated health benefits (Williams *et al*, 1991). Furthermore, physical educators have been aware for some time of their responsibility to ensure that children receive adequate levels of exercise at least within the school environment (Eston, 1984; Eston & Williams, 1986).

On account of its unequivocal success with adults, the 6-20 Rating of Perceived Exertion (RPE) scale (Borg, 1985) was initially employed to assess whether children were also able to *estimate* correctly varying levels of exercise intensity (Bar-Or, 1977; Miyashita *et al,* 1986; Ward *et al*, 1986). The findings from these studies led paediatric exercise scientists to contend that the RPE scale (Figure 2) was indeed equally valid as an indicator of exercise effort in children (above 9 years old) and adults (Bar-Or & Ward, 1989). Consequently, RPE has been used in the four studies published prior to 1995 to examine whether children could extend their understanding of effort perception by using it to regulate accurately exercise intensity (Ward & Bar-Or, 1987; Ward & Bar-Or, 1990; Ward *et al*, 1991; Williams *et al*, 1991).

Despite the above studies reporting some success with Borg's scale, an alternative scale, deemed to be more appropriate for children in terms of its language and numerical scale, has been developed recently (see Figure 3). The Children's Effort Rating Table (CERT), devised by Williams *et al.* (1994), has so far been found to be a more valid indicator of exercise effort than RPE among children aged 9 to 10 years (Study 2), and on the basis of findings from Study 1, is potentially a reliable tool for effort regulation. The purpose of the present study was to investigate the validity and reliability of both scales, CERT and RPE, as tools for regulating exercise intensity among children.

Method

Subjects

Data were initially collected on 70 children (28 boys and 42 girls), ages 9 to 10 years, representing 96% of Year 5 (4th Grade) students of a primary school in Chester, England. Each child previously had been randomly assigned to either a CERT or RPE group, and taken part in two repeat cycle ergometer *estimation* protocols. Full subject details have been described in the Methods section of Study 2. Importantly, the biometric data (height, weight, age, and estimated physical work capacity) of the two groups did not differ significantly ($p > .05$).

Procedure

Subjects were required to regulate their power outputs[1] (exercise intensities) to match a range of four scale-specific effort rating levels (randomly ordered for each child); 8, 12, 15, and 18 (RPE group), and 3, 5, 7, and 9 (CERT group). This *production* trial (P1) began with a 4 min, 25 W warm-up and continued with each child instructing the experimenter to adjust the cycling resistance in accordance with the specified perceived levels. The children were allowed 2 min to do this, before cycling for a further 2 min at the chosen intensities. The resulting objective indicators of effort, heart rate and power output, were recorded during the final 15 s of at each CERT or RPE level. A rest period of 2 min was allowed between each intensity level. One week later (same day and time) each child received a repeat trial (P2) in order to obtain a measure of the consistency of effort production with each scale.

Data Analysis

Mixed factorial ANOVAs (exercise levels by trials by sex) with repeated measures on the first two factors were applied to each group, using the MANOVA procedure in SPSS for Windows (SPSS, 1994). An advantage of this particular procedure for repeated measures designs is that the important

[1]At the onset of each production trial the children in each group were given instructions regarding what was required of them. These instructions were not read verbatim, but were relayed verbally by the experimenter in the most appropriate manner. Whilst the exact wording may have varied from child to child, the essence of the instructions was uniform.

pre-requisite of sphericity (Schutz & Gessaroli, 1987) is not jeopardized. In order to reduce the risk of inflating the experimenterwise error associated with analysing more than one dependent variable belonging to the same data set (heart rate and power output), the α-level was set at .025. Post-hoc comparisons of specific factor levels were analysed, where appropriate, with the Tukey test.

Interclass correlations (Pearson's *r*) were calculated to assess the relationship between the measures of objective and perceived effort for each group (CERT and RPE) in trials P1 and P2. For comparative purposes, a two-tailed Fisher *z* transformation (Morehouse & Stull, 1975, pp. 203-204; Appendix D2) was employed to test whether the correlation coefficients differed significantly ($p < .05$) between groups and across trials. Intraclass correlations (Safrit & Wood, 1989, pp. 52-59) obtained from a two-way analysis of variance model (in which trial-to-trial variance was accounted for) were calculated to give a quantitative indicator of the overall test-retest reliability (P1 versus P2) of effort regulation using each scale.

Results

Descriptive statistics of heart rates and power outputs for each group by prescribed level of perceived effort, trial, and sex are presented in Table 12. These data are from the 64 subjects who ultimately took part in *both* production trials (CERT: 14 boys and 17 girls; RPE: 12 boys and 21 girls). It is notable that members of the CERT group were found to have exercised consistently harder

Table 12.
Mean (SD) Heart Rates (HR) and Power Outputs (PO) by Perceived Effort Level, Trial, and Sex.

	Level	Trial	ALL HR	ALL PO	BOYS HR	BOYS PO	GIRLS HR	GIRLS PO
CERT	3	1	167.6 (25.1)	58.9 (23.8)	164.4 (25.0)	65.9 (24.4)	170.2 (25.7)	53.1 (22.3)
		2	160.8 (24.8)	50.3 (20.8)	156.4 (19.9)	52.3 (17.1)	164.5 (28.3)	48.7 (23.8)
	5	1	176.9 (14.9)	73.8 (23.4)	173.9 (15.8)	83.8 (27.2)	179.2 (14.3)	66.4 (17.5)
		2	173.1 (20.1)	69.5 (21.0)	171.6 (18.0)	75.4 (21.3)	174.2 (21.9)	65.1 (20.2)
	7	1	185.8 (17.6)	86.1 (24.0)	182.1 (13.6)	92.4 (25.2)	188.8 (20.3)	80.9 (22.4)
		2	178.3 (16.1)	77.2 (21.0)	180.3 (13.6)	89.6 (22.8)	176.6 (18.1)	67.0 (12.4)
	9	1	188.8 (13.3)	100.9 (20.9)	185.5 (10.9)	109.6 (19.4)	190.5 (14.5)	96.2 (20.7)
		2	185.1 (11.4)	94.0 (24.3)	187.7 (8.9)	116.1 (25.4)	183.9 (12.4)	82.3 (13.4)
	Overall	1	179.1 (20.1)	78.4 (27.3)	175.4 (19.0)	85.3 (28.2)	181.8 (20.5)	73.3 (25.7)
		2	173.6 (20.8)	71.3 (26.1)	172.3 (19.5)	79.4 (29.9)	174.5 (21.9)	65.3 (21.3)
RPE	8	1	148.4 (19.9)	39.7 (17.5)	147.8 (14.5)	42.7 (16.9)	148.7 (22.8)	38.0 (18.0)
		2	142.6 (18.5)	32.2 (14.0)	145.8 (18.4)	38.4 (16.4)	140.7 (18.7)	28.7 (11.4)
	12	1	166.7 (19.3)	61.2 (22.1)	166.8 (16.5)	67.9 (21.2)	166.6 (21.1)	57.4 (22.3)
		2	160.0 (22.0)	45.2 (20.7)	166.8 (15.9)	54.6 (22.4)	156.1 (24.3)	39.9 (18.0)
	15	1	178.9 (18.7)	79.5 (27.6)	179.9 (10.4)	82.2 (31.0)	178.3 (22.3)	78.0 (26.2)
		2	172.5 (15.6)	70.5 (17.0)	175.1 (11.9)	72.1 (13.2)	170.9 (17.5)	69.6 (19.2)
	18	1	183.4 (20.6)	85.0 (23.4)	186.2 (14.3)	92.3 (22.4)	181.6 (24.2)	79.9 (23.6)
		2	179.9 (19.8)	86.7 (24.2)	186.2 (10.5)	96.1 (17.6)	175.5 (23.6)	80.3 (26.5)
	Overall	1	168.7 (23.6)	65.5 (28.8)	169.8 (20.2)	70.8 (29.4)	168.0 (25.6)	62.3 (28.1)
		2	163.0 (23.5)	57.3 (28.2)	168.1 (20.5)	64.6 (27.4)	159.9 (24.7)	53.0 (27.9)

throughout both production trials than those in the RPE group, averaging approximately 10 bpm and 13 watts higher in P1 and P2.

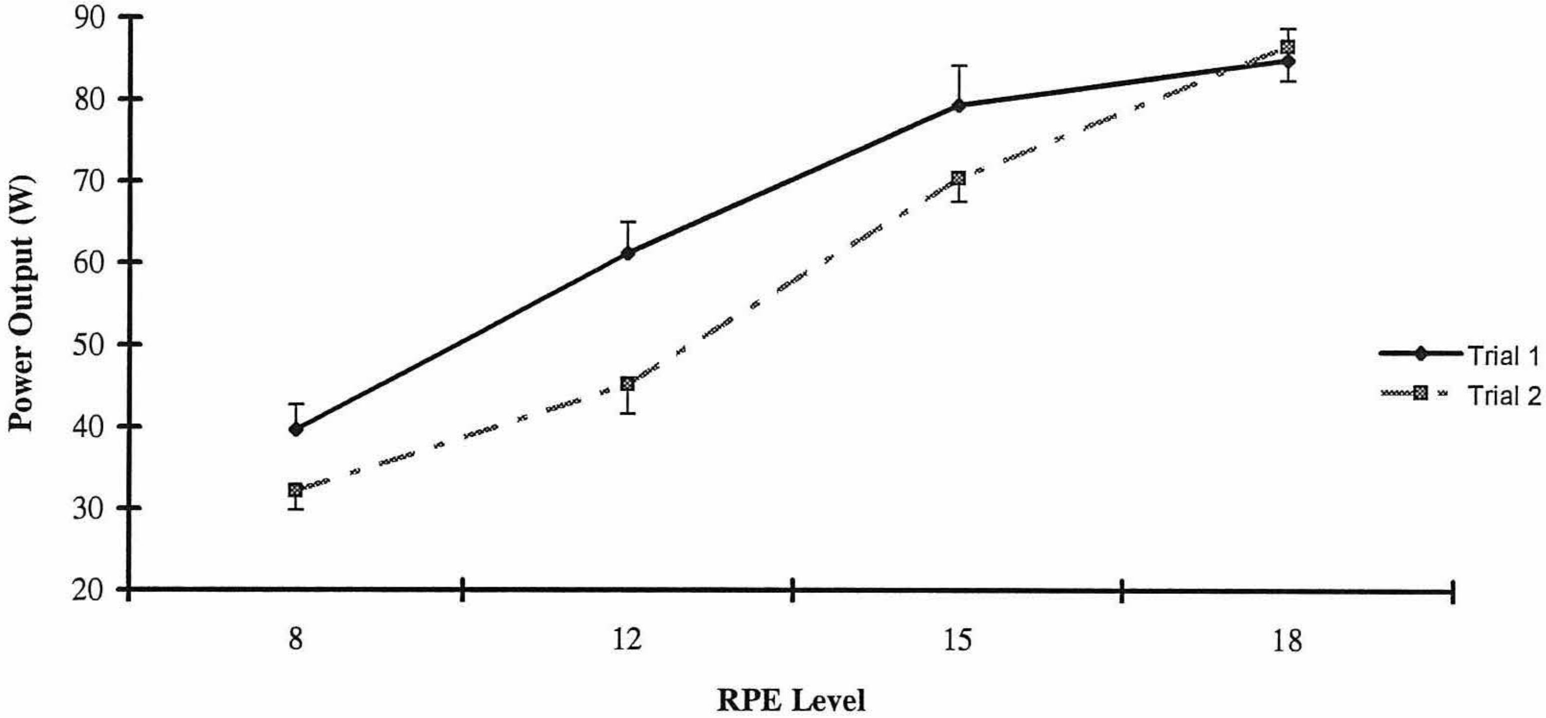
For the *heart rate* data of both groups (see Appendix C1), significant ($p < .01$) main effects were observed for perceived effort levels; $F(3, 90) = 35.3$ (CERT), $F(3, 96) = 83.6$ (RPE). Post-hoc analyses identified the differences between each successive CERT level to be significant up to CERT 7 ($T = 6.8$, $p < .05$), but not between CERT 7 and 9 (see Appendix C2). This pattern was also revealed in the RPE group ($T = 6.4$, $p < .05$), with the difference between the two highest levels (RPE 15 and 18) not reaching significance. For the CERT group, the trials main effect (represented by a mean reduction of 5.5 bpm between P1 and P2) was significant; $F(1, 30) = 6.0$, $p < .05$).

For the *power output* data of the RPE group, the 2-factor trials by levels interaction (Figure 7) was significant ($F(3, 96) = 4.2$, $p < .01$), as were the simple main effects for trials and levels (both groups; $p < .04$). In the latter, the Tukey test showed the differences between all successive CERT levels ($T = 10.7$, $p < .01$) and RPE levels ($T = 10.1$, $p < .01$) to be significant. For the CERT group only, the main effect for sex was also significant; $F(1, 30) = 15.7$, $p < .01$), with boys' mean values being higher than girls'.

Validity

Prescribed effort levels were positively and significantly ($p < .01$) correlated with the heart rate and power output values produced in both

Figure 7.
Interaction of Trials by RPE Levels on Power Output



production trials (Table 13). Correlations for the RPE scale were generally larger than those for the CERT scale, though in no case significantly so ($p > .05$), and more variable between trials. Likewise, boys' correlations were stronger than girls', particularly the heart rate correlations in both the RPE group (P1 and P2) and the CERT group (P2).

Reliability

Over the two production trials mean heart rates and power outputs were less than 6 bpm and 8 watts apart, respectively, in both groups. Such 'consistency' is better qualified by the intraclass correlation coefficients in Table 14, which reveal acceptable *overall* test-retest reliabilities for both groups. However, at the lower prescribed exercise levels, the coefficients suggest a considerable degree of trial-to-trial variability, particularly with regard to power output.

Discussion

The present study has shown that children are able, to some extent, to use their perceptions of exercise effort to regulate the magnitude of their exercise output during cycle ergometry. This finding is independent of the type of effort rating scale used (CERT or RPE), and is generally reliable on a test-retest basis. However, the data are not so convincing to suggest that either scale could readily be used in a practical sense, such as the delivery of

Table 13.
Pearson Correlations[1] Between Perceived Effort Ratings and Objective measures of Exercise Intensity.

RELATIONSHIP	EFFORT RATING SCALE: RPE P1	RPE P2	CERT P1	CERT P2
ALL				
Perceived Effort and Heart Rate	0.54	0.61	0.47	0.47
Perceived Effort and Power Output	0.59	0.75*	0.61	0.65
BOYS				
Perceived Effort and Heart Rate	0.73	0.73	0.50	0.61
Perceived Effort and Power Output	0.65	0.78*	0.59	0.75
GIRLS				
Perceived Effort and Heart Rate	0.47	0.57	0.45	0.37
Perceived Effort and Power Output	0.57	0.75	0.66	0.63

[1] All correlations significant ($p < 0.001$)
* Significantly different ($p < 0.05$) to corresponding P1 coefficient.

Table 14.
Reliability (Intraclass *R*) of P1 Versus P2 Heart Rate (HR) and Power Output (PO) Data.

Intensity Level	ALL HR	ALL PO	BOYS HR	BOYS PO	GIRLS HR	GIRLS PO
CERT						
Overall	0.70	0.74	0.72	0.76	0.69	0.64
3	0.63	0.10	0.63	0.08	0.58	-0.06*
5	0.47	0.33	0.28	0.02	0.61	0.63
7	0.49	0.63	0.42	0.74	0.54	0.33
9	0.82	0.73	0.83	0.84	0.74	0.29
RPE						
Overall	0.77	0.76	0.81	0.82	0.76	0.71
8	0.49	0.08	0.71	0.56	0.39	-0.33*
12	0.48	0.21	-0.06*	0.53	0.52	-0.37*
15	0.58	0.48	0.34	0.62	0.60	0.40
18	0.85	0.75	0.86	0.53	0.83	0.76

* Coefficients are negative because MS_B is less than MS_W.

appropriate levels of exercise by physical educators.

The analysis of variance showed that increases in prescribed RPE levels yielded increases in physiological response, thus supporting the three previous investigations using this scale (during cycle ergometry) among children aged 8-15 years (Ward & Bar-Or, 1990; Ward *et al*, 1991; Williams *et al*, 1991). Likewise, the present findings involving the CERT scale are in line with those reported in Study 1 involving 8-11 year-olds, also during cycling. Conversely, when other modes of exercise have been adopted, such as track walking or running (Ward and Bar-Or, 1990; Ward *et al*, 1991), or bench-stepping (Williams *et al*, 1994), neither scale has been found to be as successful in regulating physiological responses. For example, when required to walk or run at four different RPEs (7, 10, 13, and 16) around a 400 metre oval track, the speeds and heart rates produced did not change accordingly (Ward *et al*, 1991). With the CERT, heart rates produced by bench-stepping at ratings 5 and 7 (Williams *et al*, 1994) were almost identical for groups of 6 to 7 and 8 to 9 year-olds (though an average difference of 8 bpm was reported for 7 to 8 year-olds). On the other hand, the external loadings selected by each group of children was consistently higher for CERT 7 than CERT 5.

The significant main effect for sex on the CERT group's power output data (boys consistently producing higher power outputs than girls at increasing effort rating levels) has not been previously reported. However, one study using the RPE scale has reported opposite, though unexplained, findings among 8 to 14 year-olds during cycle ergometry. Ward *et al.* (1991) found that girls performed three levels of RPE (RPE 10, 13, and 16) at a higher *relative* power output (percentage of estimated peak power) than boys. If the present power output data are expressed in the same way, albeit from data collected in a previous estimation trial (Study 2), then the sex difference no longer exists; $F(1, 29) = 0.1, p = .82$. Notwithstanding earlier concerns about the incompatibility of data collected from estimation and production trials, it could be argued that the boys' greater power production was associated with their superior estimated physical work capacities.

More difficult to comprehend is the trials by levels interaction within the RPE group. It can be seen (Figure 7) that mean power output values are almost identical over the two production trials at the highest RPE level, yet are quite different at the three lower levels, with no particular trend evident. Such a finding may not represent a specific quirk of the RPE scale, but simply be variability related to methodological issues. For example, an unknown amount of variability can be attributed to the children not being required to exercise at the extreme anchors of each scale (CERT 1 and 10; RPE 6 and 20). This limitation presumably restricted the development of a complete frame of reference with which each child used to guide his/her adjustments in power output. In addition, the task of prescribing the CERT and RPE groups with four equivalent effort levels was, although desirable, never likely to be achieved owing to the differences in the formats of the CERT and RPE scales. Consequently, no attempt has been made to analyse the between-groups variance in the exercise intensities produced. It was interesting, however, to note that children in the RPE group consistently exercised at lower heart rates and power outputs than their CERT counterparts. At the lowest prescribed intensities (trial P1), the CERT group exercised at 19 bpm and 19 watts higher than the RPE group. At the highest intensities, the difference was 5 bpm and 16 watts. Apparently, phrases such as, "Extremely Light" (RPE 8) and "Easy" (CERT 3) do not have the same meaning, and neither does the blank space between "Very Hard" and "Extremely Hard" (RPE 18) compared with "Very, Very Hard" (CERT 9).

More tenable comparisons of the two scales can be made in terms of their associations with the objective markers of physiological strain - heart rate and power output. The magnitude of the Pearson correlations for the RPE scale is modest in P1, and improved in P2 (especially for the RPE-power output relationship). Nonetheless, the amount of unexplained variance (indicated by r^2) is substantial (44%) and implies that the scale cannot be used to regulate accurately exercise intensity. This interpretation also applies to CERT, whose correlations are more stable over the two trials, but are noticeably (though not significantly) smaller than those of the RPE scale.

Common to both scales is the stronger correlation of effort ratings with power output than with heart rate. This latter finding has been reported earlier in the two estimation studies (Studies 1 and 2), but not production. In fact, since the focus of all prior published studies using CERT or RPE has been on the so-called 'accuracy' of effort production against expected values, no inter-study contrasts can be made. However, it is reasonable to assume that when a person is requested to exercise at a 'harder', or 'less hard' perceived level than the previous one, then he/she will request a corresponding change to the applied loading. It seems that the heart rate response to such a change is less uniform, possibly as a consequence of the exercise protocol used.

Firstly, the power outputs chosen by the children at each of the four stages were established (and recorded) after two minutes of adjustment, whereas the heart rates were not recorded until almost two minutes later. Whether these values remained stable, suggesting a steady-state response, was not monitored. It is therefore possible that in some cases, and probably at the higher prescribed levels, this extra effort may have induced seemingly disproportionate increases in heart rates. Secondly, the duration of the trials (16 minutes) may in itself have been very demanding for the children, many of whom might not have been accustomed to such sustained endeavour. How this fact might explain the above is confounded by the order of the prescribed intensities having been random, whereby some children would have experienced the hardest levels first, some children the easiest, and some alternating between the two. Furthermore, given that the rest period between each level was standardised at 2 minutes, any individual differences in recovery rate could not be accounted for, adding to the variability observed. Such methodological limitations as these represent a challenge for this area of research and will be addressed in the next study (Study 4).

The overall *repeatabilities* of the heart rates and power outputs were found to be similar, and, as above, were slightly higher in the RPE group than the CERT group. However, intraclass *R*s of 0.77 (heart rate) and 0.76 (power output) for the RPE group are still indicative of within-subjects (trial-to-trial) variability, which is supported by the observed reductions in mean values of 4.7 bpm and 8.2 watts (Table 12). Comparable trial-to-trial changes are also evident for the CERT group.

The low reliability correlations calculated for each CERT effort rating level are considerably less impressive than those reported in Study 1 (intraclass *R*s of 0.95, 0.97, and 0.91 for power output during cycling at CERT 5, 7, and 9, respectively). Since the protocol was very similar to the present study, it is feasible that the larger age range (8 to 11 years) of the children in the pilot study could explain the larger correlations. This is because the size of *R* is affected not only by true within-subjects variability (M_{SW}), but by the extent of the *between-subjects* (M_{SB}) variability (Vincent, 1994). It is likely that the M_{SB} variance among the 8 to 11 year-olds in the pilot study was greater than that observed among the 9 to 10 year-olds in the present study. To elucidate, if the children elicited similar physiological responses at a given prescribed effort level, say CERT 3 (or RPE 8), M_{SB} would be small, meaning that the M_{SW} has a strong influence on the calculation of *R*. Unless M_{SW} was almost zero, *R* would be low.

That the highest effort levels (CERT 9 and RPE 18) tended to yield the highest *R*s, suggests that at these intensities the M_{SB} was large, and/or the M_{SW} was small.

Of course, it simply could be supposed that the children displayed genuine unreliability in all but the highest prescribed exercise levels. An explanation for this may rest with a number of factors, not least their levels of motivation and perceptual development. With regard to motivation, it was noted during testing that some of the children (in both groups) needed considerable verbal encouragement to complete both production trials according to the set protocol. Having been exercised twice before in the preceding two weeks (estimation trials), it was apparent that such children had not really enjoyed their experiences and were not too keen to undergo further discomfort. Researchers should be sensitive to the fact that repeated measurements on children in exercise situations may threaten the validity of their studies.

STUDY 4

The Effect of Discontinuous and Continuous Testing Protocols on Effort Perception in Children[5]

[5] The content of this chapter appeared in a paper by Lamb, K.L., Eston, R.G. & Trask, S. (1997). In N.Armstrong, B.Kirby, and J. Welsman (Eds.), *Children and Exercise XIX: Pediatric Work Physiology* (pp. 255-364). London: E. & F.N. Spon.

Introduction

Interest in the concept and application of exercise-related effort perception amongst children has been obvious, yet rather erratic over the past 25 years. Whereas research with adults has yielded an abundance of published articles on the application of effort perception in an exercise environment, particularly with regard to the use of the Rating of Perceived Exertion (RPE) scale (Borg, 1985), relatively little sustained interest has been devoted to paediatric exercise. Within the last decade, however, concerns over the physical activity and fitness levels of children have provided the motivation for researchers to explore this area with more purpose. An obvious sign of this progress is the appearance of a child-specific perceived exertion scale, the Children's Effort Rating Table (Williams *et al*, 1994), and reports on its validation (Studies 1-3 of this thesis).

Most research to date has been fairly small-scale and conducted in a "laboratory" setting, partly because of the experimental control this allows, but primarily due to its practicality. As a consequence, the potential value of perceived exertion to physical educators and health promoters remains to be realized. Over 30 published studies have focused on how well children can use their perceptions to estimate and/or regulate exercise intensity, as reflected typically by recordings of heart rate and power output (or speed), during cycle or treadmill ergometry. Yet, there appears to have been little or no attempt to standardise fundamental aspects of methodology (see Study 3), such as the structure of the exercise protocols, making it difficult to synthesise and interpret findings. For instance, when children are requested to regulate their exercise output to match experimenter-prescribed effort rating levels, their selection of loadings (resistances) and corresponding 'accuracy' is logically going to be influenced by factors such as the progression of the trial (continuous or discontinuous with rest periods), and the order of load presentation (incremental or random). No study so far has attended to these concerns. In addition, it has always been assumed that the heart rates recorded for each intensity level (usually after 4 min) reflect steady-state conditions. This may not be the case, especially during the highest prescribed levels (Study 3). Accordingly, the purpose of this investigation was to examine the effects of these protocol-related issues on children's exercise effort regulation during cycle ergometry.

Method

Subjects

Data were collected on 66 Year 5 (4th Grade) children of a primary school in Chester, England. Subject characteristics are shown in Table 15. All subjects provided parental informed consent to participate in the study and were free of known contra-indications prior to testing. Ethical approval was granted by the University College Chester Ethics Committee.

Procedure

Three weeks in advance of the testing, the children were introduced to the Children's Effort Rating Table (CERT) and the testing equipment by the experimenter. The CERT (Figure 3, Williams *et al*, 1994) is deemed to be more appropriate for children in terms of its language and numerical range, and has been offered as an alternative to the traditional RPE scale. Following an explanation of how the verbal expressions for the CERT should be interpreted

Table 15.
Subject Characteristics (means ± S.D.)

Group	n	Age (yr)	Stature (m)	Mass (kg)
Continuous				
All	33	10.30 (0.30)	1.43 (0.05)	35.08 (6.43)
Boys	21	10.29 (0.30)	1.44 (0.05)	36.10 (6.06)
Girls	12	10.33 (0.32)	1.42 (0.05)	33.29 (4.03)
Discontinuous				
All	33	10.29 (0.33)	1.45 (0.07)	36.94 (6.74)
Boys	20	10.30 (0.35)	1.46 (0.06)	38.38 (7.80)
Girls	13	10.30 (0.32)	1.44 (0.09)	34.73 (4.03)

in numerical form, every child was given a copy of the table to keep and study.

The children were then randomly assigned to either a continuous protocol (CP) or a discontinuous protocol (DP) group (see Table 15), and testing subsequently took place on the school premises over a two-week period. All exercise tests utilised a mechanically braked (basket-loading) cycle ergometer (Monark 814, Sweden) fitted with wooden pedal blocks (to accommodate the shortest children) and a radio telemetry system for monitoring heart rates (Sports Tester PE4000, Polar Electronics, Finland).

Immediately prior to exercise, subjects were re-introduced to CERT and given standard instructions concerning its use and the purpose of the test. Common to both groups was the requirement to regulate exercise intensity to match a range of four effort rating levels (randomly presented for each child); 3, 5, 7, and 9. Furthermore, each so-called *production* trial began with a 3 min 25 W warm-up, followed by a 3 min rest, and continued with each child instructing the experimenter to adjust the cycling resistance (to add or subtract weights) in accordance with the specified perceived levels. A 'shield' was in place throughout testing to hide from view the weights being applied by the experimenter.

The children were allowed 2 min to settle on an appropriate resistance, before cycling for a further 1 min at the chosen intensities. These intensities were recorded as power outputs, expressed in watts. Heart rates were also recorded after 2 min, and again at the end of the 3rd. For subjects allocated to the discontinuous group, each exercise bout was interspersed with a 3 min rest period.

Data Analysis

Exercise responses was analyzed with mixed factorial ANOVAs (Group by Levels by Time by Sex) with repeated measures on the second and third factors, using the MANOVA procedure in SPSS for Windows (SPSS, 1994). This procedure for repeated measures designs does not jeopardise the important pre-requisite of sphericity (Schutz & Gessaroli, 1987). Alpha was set at .025 in order to reduce the risk of inflating the experimenterwise error associated with analysing more than one dependent variable (heart rate and power output) belonging to the same data set. Post-hoc comparisons of specific factor levels were analyzed, where appropriate, with the Tukey test.

Additionally, and in accordance with previous research, interclass correlations (Pearson's r) were calculated to assess the relationship between the measures of objective and perceived effort for each group (CP and DP). A two-tailed Fisher z transformation (Morehouse & Stull, 1975; see Appendix D2) was employed to test whether the correlation coefficients differed significantly ($p <$ .05) between groups and across time periods.

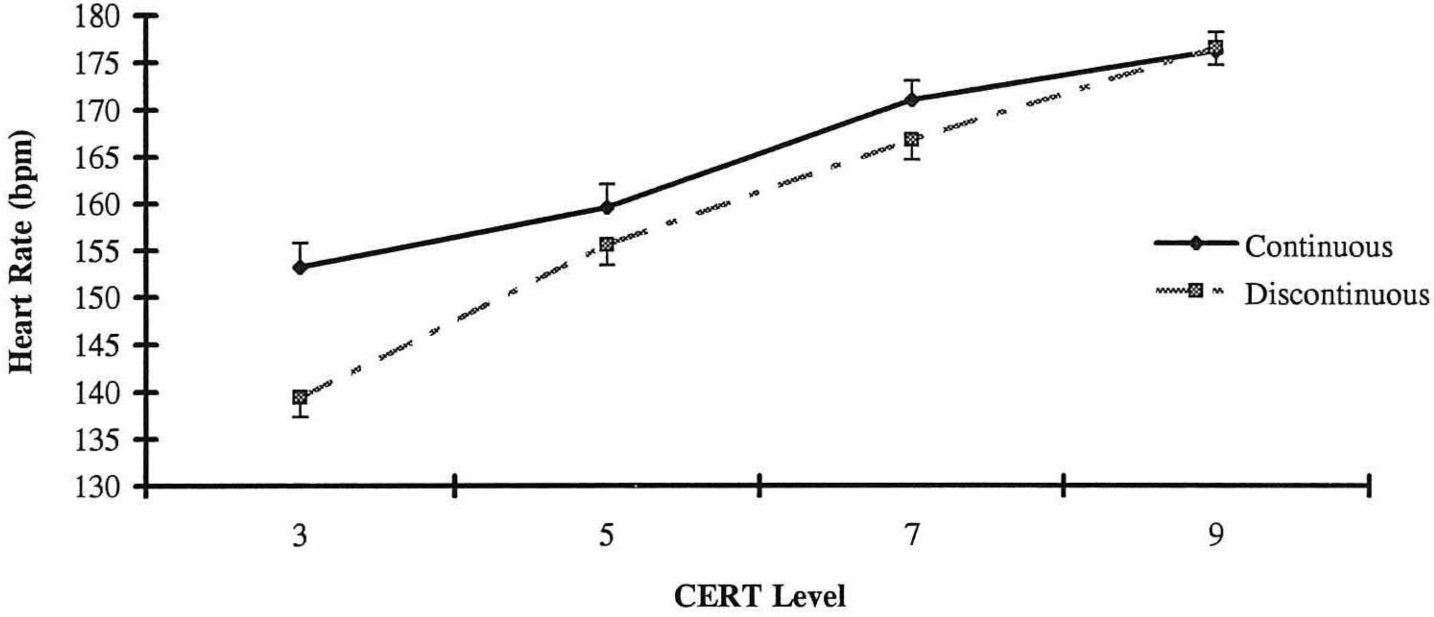
Results

For *heart rates*, mean values were consistently lower across the exercise levels (see Table 16), particularly those recorded after 2 min, though the main effect for Group was not significant ($F_{1,62} = 3.1$; $p < .05$; Appendix D1). Post-hoc analysis of the significant Group by Levels interaction ($F_{3,186} = 3.9$; $p <$ .01) identified the heart rates at CERT 3 to be significantly different ($T = 12.7$; p $< .05$), but not elsewhere (see Figure 8).

Table 16.
Mean (±SD) Heart Rates (bpm) After 2 Min and 3 Min by Perceived Effort Level and Sex.

Group	Level	ALL 2 min	ALL 3 min	BOYS 2 min	BOYS 3 min	GIRLS 2 min	GIRLS 3 min
CONTINUOUS	3	152.5 (21.4)	153.8 (20.9)	148.9 (16.5)	150.6 (16.1)	158.8 (27.7)	159.5 (27.3)
	5	157.8 (20.6)	161.4 (20.5)	155.9 (20.2)	158.1 (19.5)	161.2 (21.9)	167.2 (21.7)
	7	168.4 (16.5)	173.8 (16.0)	164.6 (14.4)	167.9 (14.3)	175.1 (18.3)	184.0 (13.7)
	9	174.0 (16.3)	178.5 (16.9)	171.4 (14.3)	175.2 (16.1)	178.7 (19.2)	184.2 (17.5)
	Overall	163.2 (20.5)	166.8 (20.9)	160.2 (18.4)	162.9 (18.8))	168.4 (23.1)	173.7 (22.8)
DISCONTINUOUS	3	137.9 (15.8)	141.0 (18.2)	135.0 (14.3)	138.0 (16.4)	142.5 (17.4)	145.7 (20.5)
	5	152.1 (15.7)	159.2 (18.7)	152.0 (15.5)	158.5 (16.7)	152.3 (16.7)	160.2 (22.1)
	7	162.8 (16.2)	170.8 (16.6)	160.3 (14.5)	168.6 (15.7)	166.8 (18.4)	174.2 (18.1)
	9	171.4 (13.6)	181.8 (14.2)	168.6 (13.4)	178.5 (13.9)	175.7 (13.3)	186.9 (13.7)
	Overall	156.1 (19.7)	163.2 (22.6)	154.0 (18.9)	160.9 (21.5)	159.3 (20.6)	166.8 (24.0)

Figure 8.
Interaction of Group by CERT Levels on Heart Rate



Values recorded after 2 min exercise were found to be generally lower than those after 3 min ($F_{1,62}$ = 17.4; $p < .01$), but were also influenced by Group ($F_{1,62}$ = 13.8; $p < .01$) and Levels ($F_{3,186}$ = 12.8; $p < .01$; Figure 9). Lastly, post-hoc analysis of the highly significant ($p < .01$) Levels factor confirmed that heart rates increased stepwise with increases in prescribed CERT ratings (T = 6.9; $p < .05$).

For *power output* data (see Table 17), the most notable finding was a Group by Sex interaction ($F_{1,62}$ = 8.2; $p < .01$), accounted for principally by a mean difference of 12.4 W between the girls in the two groups (T = 10.7; $p < .05$). Likewise, the significant Group by Levels interaction ($F_{3,186}$) = 3.9; $p < .01$) was mainly due to a 13.5 W between-groups difference in power outputs at the lightest CERT level (T = 13.4; $p < .01$).

Correlation coefficients between heart rates and CERT levels were consistently and significantly ($p < .05$) larger in the DP group than the CP group (Table 18), whereas those for power output and CERT levels only differed significantly for boys.

Discussion

The present data have demonstrated that children's use of perceived effort to regulate exercise intensity during cycle ergometry is, to some extent, protocol-dependent. Both the nature of the exercise protocol and the point at which data

Figure 9.
Interaction of Time by CERT Levels on Heart Rate

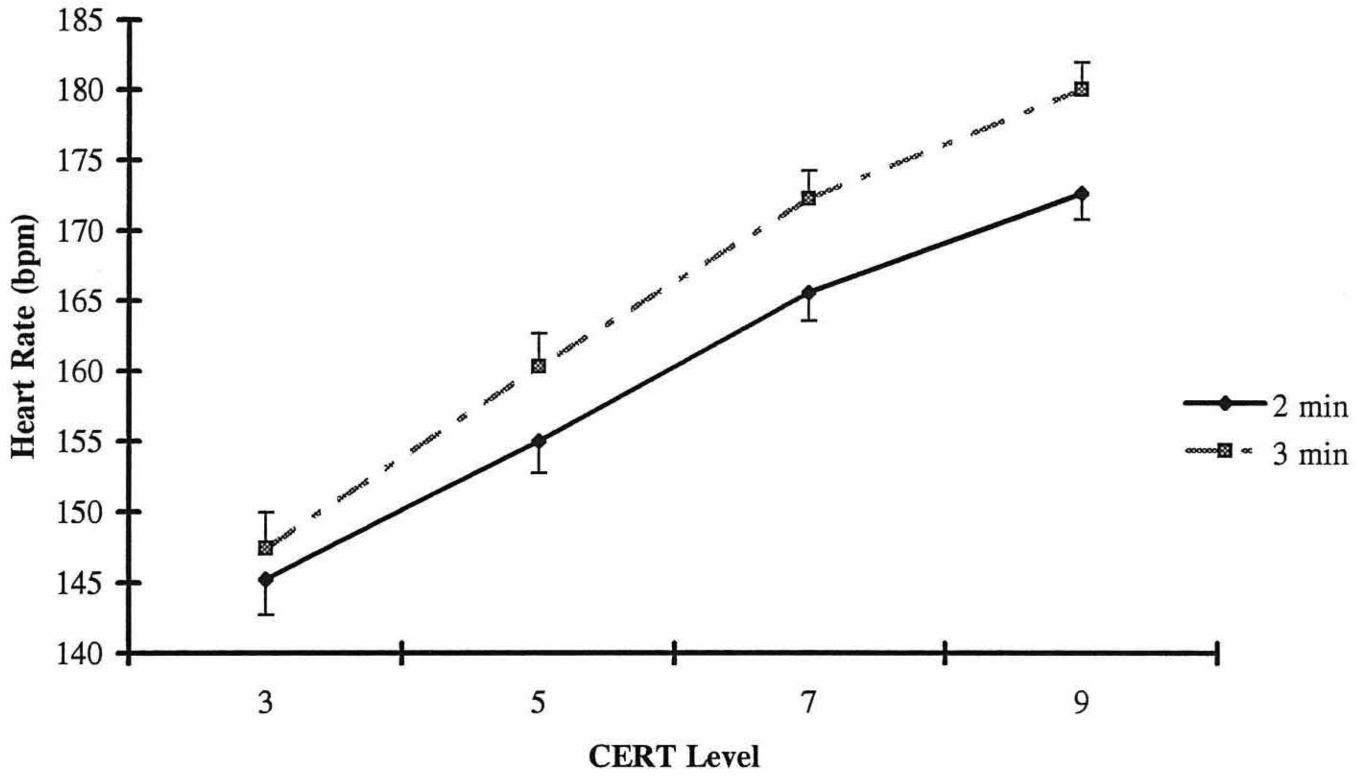


Table 17.
Mean (±SD) Power Outputs (watts) After 2 Min by Perceived Effort Level and Sex.

Group	Level	ALL	BOYS	GIRLS
CONTINUOUS	3	52.1 (16.2)	52.9 (21.3)	55.8 (17.6)
	5	68.2 (17.0)	66.4 (17.1)	71.3 (16.9)
	7	89.4 (17.7)	87.9 (20.6)	92.1 (11.2)
	9	99.6 (15.0)	99.5 (16.8)	99.6 (12.0)
	Overall	77.2 (24.6)	76.0 (25.8)	79.7 (22.5)
DISCONTINUOUS	3	38.6 (14.4)	41.5 (14.5)	34.2 (13.5)
	5	70.5 (20.8)	75.0 (15.1)	63.5 (26.6)
	7	89.4 (21.4)	94.3 (16.2)	81.9 (26.5)
	9	102.7 (20.2)	111.3 (15.5)	89.6 (13.3)
	Overall	75.3 (30.9)	80.5 (30.1)	67.3 (30.5)

Table 18.
Pearson Correlations[1] Between CERT Ratings and Objective Measures of Exercise Intensity.

RELATIONSHIP	GROUP			
	Continuous		**Discontinuous**	
	2 min	3 min	2 min	3 min
ALL				
Perceived Effort and Heart Rate	0.41	0.46	0.63*	0.66*
Perceived Effort and Power Output	0.74	-	0.77	-
BOYS				
Perceived Effort and Heart Rate	0.47	0.50	0.65*	0.69*
Perceived Effort and Power Output	0.74	-	0.85*	-
GIRLS				
Perceived Effort and Heart Rate	0.36	0.45	0.62*	0.65
Perceived Effort and Power Output	0.76	-	0.68	-

[1] All correlations significant ($p < 0.01$)
* Significantly different ($p < 0.05$) to corresponding Continuous Group coefficient.

are recorded have an important bearing on the heart rate responses at specific perceived effort levels. In particular, the provision of recovery periods between exercise bouts seems to enhance their ability to utilize a scale such as the CERT, as does the duration of each bout. However, on a cautionary note, in studies of this kind it remains extremely difficult to separate the variance due to these factors from that due simply to individual differences.

The findings that the children in this study had some success in using their perceptions of effort to increase or decrease exercise intensity corroborate existing knowledge acquired using discontinuous (Ward & Bar-Or, 1990; Ward *et al* , 1991), and continuous (Williams *et al*, 1991) cycling. No previous effort production studies have compared these two conditions. Of note in the present study, therefore, are the significant interactions involving the Group factor. Whilst the lower heart rates (and power outputs) recorded for CERT 3 in the DP group are striking, being isolated they are difficult to interpret. This is especially true since loadings were randomly applied for each group. Also, that the mean heart rates recorded for the CP (176.3 bpm) and DP (176.6 bpm) groups were no different at the hardest prescribed level (CERT 9), implies nothing was gained by the rest periods in the DP. Similarly, an explanation for the Group by Time interaction in which only the 2 min heart rates were disproportionately lower in the DP group than the CP group, is not obvious. The simple main effect of Time on heart rates, is, however, probably due to the non-attainment of steady-state conditions after 2 min of exercise. Whether this status was reached even after 3 min is not known.

Clearer evidence for the impact of the type of protocol on effort regulation is provided by the bivariate correlational analysis. The discontinuous protocol produced a consistently stronger relationship between perceived effort and heart rate levels, hinting that recovery periods may assist children in their perceptions of physiological strain. Conversely, the lower correlations in the CP group may simply reflect an inaccuracy of perception due as much to fatigue as error of judgement.

In conclusion, future investigations into children's effort perception should not disregard the manner in which the exercise stimulus is applied, nor indeed, the duration of the stimulus, since these two factors seem to modify significantly the outcome measures.

CONCLUSIONS

The research presented in this thesis has shed new light on the study of exercise-related effort perception in children. In so doing, it will hopefully serve to quell the small tide of inappropriate exploration which has been gathering in this field over the past years, and stimulate more structured and varied programmes of investigation. As it now stands, the body of knowledge is not definitive; children seem to have some grasp of the concept of effort perception and can best relate to the CERT as a method of expressing such understanding, yet the likelihood that they can extend this process to particular situations requiring self-regulation of exercise, is still unknown.

The current research has also highlighted that the manner of data analysis can notably influence the interpretation of how proficient children are with their perceptions of effort. Data from Studies 2 and 3 confirmed previous suspicions that the associations between effort ratings and objective indicators were probably spurious (actually too high) on account of the way in which the relevant statistic - the correlation coefficient - was calculated. Moreover, it emerged that the "established" custom of comparing effort perception responses during estimation and production trials is wholly unjustified and unnecessary. Accordingly, future investigators should not take for granted this essential component of the research process

With particular regard to the performance of the relatively new CERT scale, the existing research is equivocal. On the one hand, data now exists to vindicate its development as a more suitable tool for children than the RPE scale. On the other hand, such data does not imply that further progress in this field is unwarranted. Data from Studies 2 and 3 raise questions about the reliability of the CERT, especially when applied in its production mode. Without being reliable, no measurement tool can be valid. Furthermore, the situation here may be confounded by the recent emergence of an argument against the type of reliability analysis used in the present studies. Specifically, readers of the medical and epidemiological literature may be familiar with the assertions made by Bland and Altman (1986; 1995) regarding the assessment of reliability.

These statisticians and others (e.g. Atkinson, 1995) argue that the intraclass correlation (the statistic used in this thesis) is too sensitive to the range of measurements in a sample (between subjects variance), that is, if the sample is homogeneous in the variable of interest, the intraclass correlation is likely to be small, even if the trial-to-trial variability is low. Instead, Bland and Altman offer as a better method their relatively simple technique of calculating the 95% *limits of agreement* between repeated measurements, which are unaffected by sample heterogeneity and allow the investigator to interpret the *practical* significance of any differences observed.

A belated application of this novel technique to the data of Studies 1-3 (see Appendix E1) revealed findings which tend to reinforce those conclusions already formed on the basis of the intraclass correlation statistic. This is not unexpected since the data in each study were indeed found to be heterogeneous on the variables of interest, and therefore suitable for the initial form of analysis. However, the new method does provide a more qualitative appraisal of reliability. For example, limits of agreement figures for CERT in estimation mode (Study 2) reveal an overall mean test-retest difference (bias) of 0.3 units, which is seemingly quite impressive. However, the +/- 2.6 limits of agreement indicate that some children had estimations of effort perception that varied by almost 3 categories between test and retest. In semantic terms, this could be the difference between the exercise effort being rated as "Getting Quite Hard" or "Very, Very Hard". Similarly, with regard to using CERT in production mode (Study 3), an overall mean test-retest bias of 7.1 watts (power output) and 5.4 b/min (heart rate) seems negligable, yet these biases were as high as 49 watts and 38 b/min or more, respectively, for some of the children.

Notwithstanding the above comments, the preceding studies have shown that:

(i) when asked to express their overall feelings of exertion during incremental cycling exercise, the findings from Studies 1 and 2 represent the most credible to date. The children's responses (with either the CERT or its predecessor, the RPE scale) to the physiological strain imposed on them yielded correlations (0.69-0.80) that, whilst not trivial, indicate considerable inter-individual variability exists. Nevertheless, the validity of the CERT is superior to that of RPE scale and in the short-term deserves to be adopted in future studies of this kind. The overall repeatability of the above responses is satisfactory, though apparently influenced by the intensity of exercise imposed (Study 2).

(ii) when asked to manipulate the exercise loadings to match specific perceived exertion levels, the objective-subjective intensity associations are, whilst significant, less impressive than in the estimation mode. In this production mode, the children seem not to be able to use the CERT any better than the RPE scale, and in both cases, are not consistent on a trial-to-trial basis. Findings such as these suggest the need for further investigation into scale development and/or the factors influencing children's effort sense.

(iii) The nature of the exercise protocols can influence the responses given and subsequent interpretations. More precisely, children are better able to use their perceptions of effort (via CERT) in production mode, when the exercise is discontinuous (with rest periods), rather than continuous. This finding emphasizes the need for appropriate and standardised methodologies to be used in future investigations.

It is transparent from being involved with research of this kind that the scope for fundamental research is considerable. It seems that much exploration and knowledge accumulation has been 'by-passed' in the rapid pursuit of an end-product. Research with children *has* produced a large body of knowledge on the way in which they perceive and understand their environment, and how such awareness develops from a young age. Past research has focused on how they manage perceptual continua, such as vision, touch, size, position, and orientation, and the role of issues such as *absolute and relative codes*, and *external frames of reference* (Bryant, 1974). Yet, no research to date has considered children's perceptions of exercise effort in such a way. Answers to questions such as "Are children's perceptions of effort totally reliant upon their ability to register and remember relative values ('harder than' or 'easier than'), or do they have absolute values to which they can turn for reference?", would help explain the present (and previous) findings. The stage of development of effort perception is likely to have a direct bearing on the ability of children to regulate exercise effort successfully.

It is unfortunate that attempts so far to interpret the findings of the previously published production studies have been meagre. Researchers have hinted at the importance of the extent of children's experiences of exercise on effort perception (Borg, 1977; Ward *et al*, 1991), or (in an experimenter-controlled environment) the absence of any influence on their self-regulatory systems by physically superior peers (Williams *et al*, 1991). No effort has been made to study systematically the use of effort rating scales among children, in estimation *or* production mode, and to account for variables such as habituation, sex, activity level, reading ability, exercise mode and intensity, or level of perceptual development. On the other hand, such rigour may be unnecessary for the development of a tool which possesses a high degree of external validity. What “works” in a physical education setting might be far removed from anything found to be valid in a controlled, laboratory-type environment. Indeed, continuing on this positive note, the likely consumers of this type of research - physical educators, exercise scientists, and health promoters - should not be deterred by the current findings from attempting to use items such as the CERT in the ‘field’. If it transpires that children can only satisfactorily distinguish between three levels of exertion during a physical education class, (say, “Easy”, “Hard”, and “So Hard I’m Going to Stop”), rather than being disappointed, the teacher should be encouraged to make use of this awareness.

REFERENCES

Alekseev, V.M. (1989). Correlation between heart rate and subjectively perceived exertion during muscular work. *Human Physiology*, 15, 39-44.

American College of Sports Medicine (1991). *Guidelines for exercise testing and prescription* (pp. 101-102). Philadelphia, PA: Lea & Febiger.

Arnhold, R., Ng, N. & Pechar, G. (1992). Relationship of rated perceived exertion to heart rate and workload in mentally retarded young adults. *Adapted Physical Activity Quarterly,* 9, 47-53.

Atkinson, G. (1995). A comparison of statistical methods for assessing measurement repeatability in ergonomics research. In G. Atkinson & T. Reilly (Eds.), *Sport, Leisure and Ergonomics* (pp. 218-222). London: E. & F.N. Spon.

Bar-Or, O. (1977). Age-related changes in exercise perception. In G. Borg (Ed.), *Physical Work and Effort* (pp. 255-256). Solna, Sweden: Pergamon Press.

Bar-Or, O. (1983). *Pediatric Sports Medicine for the Practitioner.* (p. 25). New York: Springer-Verlag.

Bar-Or, O. & Reed, S. (1986). Rating of perceived exertion in adolescents with neuromuscular disease. In G. Borg & D. Ottoson (Eds.), *The Perception of Exertion in Physical Work* (pp. 137-148). Basingstoke: The Macmillan Press Ltd..

Bar-Or, O. (1989). Rating of perceived exertion in children and adolescents: clinical aspects. In G. Ljunggren & S. Dornic (Eds.), *Psychophysics in Action.*(pp. 105-113). Berlin: Springer-Verlag.

Bar-Or, O. & Ward, D.S. (1989). Rating of perceived exertion in children. In O. Bar-Or (Ed.), *Advances in Pediatric Sports Sciences, vol. 3* (pp. 151-168). Champaign: Human Kinetics Books.

Bartlett, R. (1997). The use and abuse of statistics in sport and exercise sciences. *Journal of Sports Sciences*, 15, 1-2.

Bland, J.M. & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, i 307-310.

Bland, J.M. & Altman, D.G. (1995). Comparing two methods of clinical measurement: a personal history. *International Journal of Epidemiology*, 24 (Suppl. 1), S7-S14.

Borg, G. & Linderholm, H. (1967). Perceived exertion and pulse rate during graded exercise in various age groups. *Acta Medica Scandinavica*, 412, 184-200.

Borg, G. (1970). Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine,* 2, 92-98.

Borg, G.& Noble, B. (1974). Perceived exertion. *Exercise and Sports Science Reviews,* 2, 131-153.

Borg, G. (1977). *Physical Work and Effort* (p. 289). Oxford: Pergamon Press.

Borg, G. (1985). *An Introduction to Borg's RPE-Scale*. Ithaca, New York: Mouvement Publications.

Borg, G. (1986). Psychophysical studies of effort and exertion: some historical, theoretical and empirical aspects. In G. Borg & D. Ottoson (Eds.), *The Perception of Exertion in Physical Work* (pp. 3-12). London: Macmillan.

Bryant, P.E. (1974). *Perception and Understanding in Young Children*. London: Methuen & Co. Ltd.

Carton, R.L. & Rhodes, E.C. (1985). A critical review of the literature on rating scales for perceived exertion. *Sports Medicine*, 2, 198-222.

Clapp, J.F. & Little, K.D. (1994). The physiological response of instructors and participants to three aerobics regimens. *Medicine and Science in Sports and Exercise*, 26, 1041-1046.

Davies, C.T.M., Fohlin, L. & Thoren, C. (1980). Perception of exertion in anorexia nervosa patients. In K. Berg & B.O. Eriksson (Eds.), *Children and Exercise IX* (pp. 327-332). Baltimore: University Park Press.

Dunbar, C.C., Robertson, R.J., Baun, R., Blandin, M.F., Metz,K., Burdett, R. & Goss, F.L. (1992). The validity of regulating exercise intensity by ratings of perceived exertion. *Medicine and Science in Sports and Exercise*, 24, 94-99.

Dunbar, C.C., Goris, C., Michielli, D.W. & Kalinski, M.I. (1994). Accuracy and reproducibility of an exercise prescription based on ratings of perceived exertion for treadmill and cycle ergometer exercise. *Perceptual and Motor Skills*, 78, 1335-1344.

Duncan, G.E., Mahon, A.D., Gay, J.A. & Sherwood, J.J. (1996). Physiological and perceptual responses to graded treadmill and cycle exercise in male children. *Pediatric Exercise Science*, 8, 251-258.

Eakin, B.L., Finta, K.M., Serwer, G.A. *et al.* Perceived exertion and exercise intensity in children with or without structural heart defects. *Journal of Pediatrics*, 120, 90-93.

Edwards, A.L. (1976). *An Introduction to Linear Regression and Correlation.* San Francisco: W.H. Freeman & Company.

Ekblom, B. & Goldbarg, A.N. (1971). The influence of physical training and other factors in the subjective rating of perceived exertion. *Acta Physiologica Scandinavica*, 83, 399-406.

Eston, R.G. (1984). A discussion of the concepts: exercise intensity and perceived exertion with reference to the secondary school. *Physical Education Review,* 7, 19-25.

Eston, R.G. & Williams, J.G. (1986). Exercise intensity and perceived exertion in adolescent boys. *British Journal of Sports Medicine*, 20, 27-30.

Eston, R.G., Davies, B.L. & Williams, J.G. (1987). Use of perceived effort ratings to control exercise intensity in young healthy adults. *European Journal of Applied Physiology*, 56, 222-224.

Eston, R.G. & Williams, J.G. (1988). Reliability of ratings of perceived effort in the regulation of exercise intensity. *British Journal of Sports Medicine*, 2, 53-155.

Gillach, M.C., Sallis, J.F., Buono, M.J., Patterson, P. & Nader, P.R. (1989). The relationship between perceived exertion and heart rate in children and adults. *Pediatric Exercise Science,* 1, 360-368.

Glass, S.C., Knowlton, R.G. & Becque, M.D. (1992). Accuracy of RPE from graded exercise to establish exercise training intensity. *Medicine and Science in Sports and Exercise*, 24, 1303-1307.

Hagen, K.B., Vik, T., Myhr, N.E., Opsahl, P.A. & Harms-Ringdahl, K. (1993). Physical workload, perceived exertion, and output of cut wood as related to age in motor-manual cutting. *Ergonomics*, 36, 479-488.

Kahle, C., Ulmer, H.V. & Rummel, L. (1977). The reproducibility of Borg's RPE scale with female pupils from 7 to 11 years of age. *European Journal of Physiology*, 368, R26 (Abstract).

Kerlinger, F.N. & Pedazhur, E.J. (1973). *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart and Winston. Inc..

Koltyn, K.F. & Morgan, W.P. (1992).Efficacy of perceptual versus heart rate monitoring in the development of endurance. *British Journal of Sports Medicine*, 26, 132-134.

Mahon, A.D. & Marsh, M.L. (1992) Reliability of the rating of perceived exertion at ventilatory threshold in children. *International Journal of Sports Medicine*, 13, 567-571.

Mahon, A.D. & Ray, M.L. (1995). Ratings of perceived exertion at maximal exercise in children performing different graded exercise tests. *Journal of Sports Medicine and Physical Fitness* , 35, 38-42.

Marriott, H.E. & Lamb, K.L. (1996). The use of ratings of perceived exertion for regulating exercise levels in rowing ergometry. *European Journal Of Applied Physiology*, 72, 267-271.

Meyer, F., Bar-Or, O. & Wilk, B. (1995) Children's perceptual responses to ingesting drinks of different compositions during and following exercise in the heat. *International Journal of Sports Nutrition*, 5, 13-24.

Mihevic, P.M. (1981). Sensory cues for perceived exertion: a review. *Medicine and Science in Sports and Exercise*, 13, 150-163.

Miyashita, M., Onedera, K. & Tabata, I. (1986). How Borg's RPE scale has been applied to Japanese. In G. Borg & D. Ottoson (Eds.) *The Perception of Exertion in Physical Work* (pp. 27-34). Basingstoke: The Macmillan Press.

Morehouse, C.A. & Stull, G.A. (1975). *Statistical Principles and Procedures with Applications for Physical Education*. Philadelphia: Lea & Febiger.

Morgan, W.P. (1973). Psychological factors influencing perceived exertion. *Medicine and Science in Sports*, 5, 97-103.

Munro, B.H. & Page, E.B. (1993). *Statistical Methods for Health Care Research*. Philadelphia: J.B. Lippincott Company.

Myles, W.S. & Maclean, D. (1986). A comparison of response and production protocols for assessing perceived exertion. *European Journal of Applied Physiology,* 55, 585-587.

Nevill, A. (1996). Validity and measurement agreement in sports performance. *Journal of Sports Sciences*, 14, 199.

Noble, B.J. (1982). Clinical applications of perceived exertion. *Medicine and Science in Sports and Exercise*, 14, 406-411.

Noble, B.J. & Robertson, R.J. (1996). *Perceived Exertion*. Champaign: Human Kinetics.

Nystad, W., Oseid, S. & Mellbye, E.B. (1989). Physical education for asthmatic children: the relationship between changes in heart rate, perceived exertion, and motivation for participation. In S. Oseid & K. Carlsen (Eds.), *Children and Exercise XIII* (pp. 369-377). Champaign: Human Kinetics Books.

Pandolf, K.B. (1983). Advances in the study of perceived exertion. *Exercise and Sports Science Reviews*, 11, 118-138.

Parfitt, G., Eston, R.G., & Connolly, D.A. (1996). Psychological affect at different ratings of perceived exertion in high- and low-active women: implications for exercise promotion. *Journal of Sports Sciences*, 14, 42-43.

Safrit, M.J. & Wood, T.M. (1989). *Measurement Concepts in Physical Education and Exercise Science*. Champaign: Human Kinetics Books.

Schutz, R.W. & Gessaroli, M.E. (1987). The analysis of repeated measures designs involving multiple dependent variables. *Research Quarterly for Exercise and Sport*, 58 132-149.

Shephard, R.J., Kavanagh, T., Mertens, D.J. & Yacoub, M. (1996). The place of perceived exertion ratings in exercise prescription for cardiac transplant patients before and after training. *British Journal of Sports Medicine*, 30, 116-121.

Skinner, J.S., Hutsler, R., Bergsteinova, V. *et al.*(1973). The validity and reliability of a rating scale of perceived exertion. *Medicine and Science in Sports*, 5, 94-96.

Smutok, M.A., Skrinar, G.S. & Pandolf, K.B. (1980). Exercise intensity: subjective regulation by perceived exertion. *Archives of Physical Medicine and Rehabilitation,* 61, 569-574.

SPSS (1990). *SPSS/PC+ Statistics 4.0.* Chicago: SPSS Inc..

SPSS (1994). *SPSS for Windows 6.1.* Chicago: SPSS Inc..

Stamford, B.A. (1976). Validity and reliability of subjective ratings of perceived exertion during work. *Ergonomics*, 19, 53-60.

Stratton, G. & Armstrong, N. (1994). Children's use of RPE during indoor handball lessons. *Journal of Sports Sciences*, 12, 182-183.

Tolfrey, K., & Mitchell, J.(1996). Ratings of perceived exertion at standard and relative exercise intensities in prepubertal, teenage and young adult males. *Journal of Sports Sciences*, 14, 101-102.

Tones, S.S. (1995). Understanding effort perception scales. *Unpublished Manuscript.*

Ueda, T. & Kurokawa, T. (1991). Validity of heart rate and ratings of perceived exertion as indices of exercise intensity in a group of children while swimming. *European Journal of Applied Physiology*, 63, 200-204.

Ueda, T. & Kurokawa, T. (1995). Relationships between perceived exertion and physiological variables in swimming. *International Journal of Sports Medicine*, 16, 385-389.

Van den Burg, M. & Ceci, R. (1986). A comparison of psychophysical estimation and a production method in a laboratory and a field condition. In G. Borg & D. Ottoson (Eds.), *The Perception of Exertion in Physical Work* (pp. 35-46). London: Macmillan.

Van Huss, W.D., Stephens, K.E., Vogel, P. *et al.*(1986). Physiological and perceptual responses of elite age group distance runners during progressive intermittent work to exhaustion. In M. Weiss & D. Gould (Eds.), *Sport for Children and Youth* (pp. 239-246). Champaign: Human Kinetics Publishers.

Vincent, W.J. (1994). *Statistics in Kinesiology*. Champaign: Human Kinetics.

Walker, C.A.H., Lamb, K.L. & Marriott, H.E. (1996). The validity of using ratings of perceived exertion to estimate and regulate exercise intensity during stepping ergometry. *Journal of Sports Sciences*, 14, 102-103.

Ward, D.S., Blimkie, C.J.R. & Bar-Or, O. (1986). Rating of perceived exertion in obese adolescents. *Medicine and Science in Sports and Exercise*, 18 (Suppl.), S72 (Abstract).

Ward, D.S. & Bar-Or, O. (1987). Usefulness of RPE scale for exercise prescription with obese youth. *Medicine and Science in Sports and Exercise*, 19 (Suppl.), S15 (Abstract).

Ward, D.S. & Bar-Or, O. (1990). Use of the Borg scale in exercise prescription for overweight youth. *Canadian Journal of Sports Sciences*, 15, 120-125.

Ward, D.S., Jackman, J.D. & Galiano, F.J. (1991). Exercise intensity reproduction: children versus adults. *Pediatric Exercise Science*, 3: 209-218.

Ward, D.S., Bar-Or, O., Longmuir, P. & Smith, K. (1995). Use of ratings of perceived exertion (RPE) to prescribe exercise intensity for wheelchair-bound children and adults. *Pediatric Exercise Science*, 7, 94-102.

Watt, B. & Grove, R. (1993). Perceived exertion: antecedents and applications. *Sports Medicine*, 14, 225-241.

Williams, J.G. & Eston, R.G. (1989). Determination of the intensity dimension in vigorous exercise programmes with particular reference to the use of the rating of perceived exertion. *Sports Medicine*, 8, 177-189.

Williams, J.G., Eston, R.G. & Stretch, C. (1991). Use of rating of perceived exertion to control exercise intensity in children. *Pediatric Exercise Science,* 3, 1-27.

Williams, J.G., Furlong, B., Mackintosh, C. & Hockley, T-J. (1993). Rating and regulation of exercise intensity in young children. *Medicine and Science in Sports and Exercise*, 25 (Suppl.), S8 (Abstract).

Williams, J.G., Eston, R.G. & Furlong, B. (1994). CERT: a perceived exertion scale for young children. *Perceptual and Motor Skills*, 79, 1451-1458.

Williams, J.G. & Eston, R.G. (1996). Exercise intensity regulation. In R.G. Eston & T. Reilly (Eds.), *Kinanthropometry and Exercise Physiology Laboratory Manual* (pp. 229-232). London: E. & F.N. Spon.

APPENDICES

A1. REPEATED MEASURES ANOVA FOR STUDY 1 DATA (PRODUCTION TRIALS)

Heart Rate

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	9232.83	14	659.49		
SEX	3408.17	1	3408.17	5.17	.039

Tests involving 'TRIALS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1981.50	14	141.54		
TRIALS	140.17	1	140.17	.99	.337
SEX BY TRIALS	150.00	1	150.00	1.06	.321

Tests involving 'LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1132.42	28	40.44		
LEVELS	8791.58	2	4395.79	108.69	.000
SEX BY LEVELS	10.33	2	5.17	.13	.881

Tests involving 'TRIALS BY LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	528.00	28	18.86		
TRIALS BY LEVELS	4.08	2	2.04	.11	.898
SEX BY TRIALS BY LEVELS	12.25	2	6.13	.32	.725

Power Output

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	23909.90	14	1707.85		
SEX	1544.01	1	1544.01	.90	.358

Tests involving 'TRIALS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	947.40	14	67.67		
TRIALS	2.34	1	2.34	.03	.855
SEX BY TRIALS	396.09	1	396.09	5.85	.030

Tests involving 'LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	991.67	28	35.42		
LEVELS	8864.06	2	4432.03	125.14	.000
SEX BY LEVELS	69.27	2	34.64	.98	.389

Tests involving 'TRIALS BY LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	429.17	28	15.33		
TRIALS BY LEVELS	4.69	2	2.34	.15	.859
SEX BY TRIALS BY LEVELS	57.81	2	28.91	1.89	.170

A2. INTRACLASS RELIABILITY CORRELATION: WORKED EXAMPLE (STUDY 1)

INTRACLASS *R* = (MSs - MSw) / MSs

where MSs = mean squares *between* subjects,

and MSw = (SStrials + SSwithin) / (df trials + df within)

The following example is taken from the heart rate data (presented in A1, above)

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	9232.83	14	659.49		
SEX	3408.17	1	3408.17	5.17	.039

Tests involving 'TRIALS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1981.50	14	141.54		
TRIALS	140.17	1	140.17	.99	.337
SEX BY TRIALS	150.00	1	150.00	1.06	.321

MSs = 659.49

MSw = (140.17 + 1981.50) / (1 + 14)
= 2121.67 / 15
= 141.44

R = (659.49 - 141.44) / 659.49

R = <u>0.79</u>

B1. LETTER OF INFORMED CONSENT

Dear Parent/Guardian,

Overleigh St. Mary's Primary School has agreed to become involved in an important research project being supervised by Senior Lecturers in the Department of Physical Education & Sports Science at Chester College of Higher Education. The project is investigating how well children can gauge their effort during exercise. It is hoped that the findings will assist P.E. teachers in ensuring that school children receive the quality of exercise which promotes health and physical fitness.

Your child has been selected to take part in this project and I would be grateful for your permission for him/her to be involved. He/she will be exercised on a cycle for about 20 minutes on four weekly occasions during September/October. We will be measuring your child's heart rate during each test and asking him/her to tell us how hard they *feel* they are exercising. The cycling will range from "very easy" to "very hard", but it will not be exhaustive (the hardest level will be similar to that experienced playing football or netball).

Please could you complete the details requested below and return the whole form to the school as soon as possible. (Please note that P.E. kit will be required for each occasion.)

Thankyou in anticipation.

Yours sincerely,

Kevin L. Lamb
Senior Lecturer in P.E. and Sports Science, Chester College of H.E.

1. Child's Name ....
2. Age ....
3. Date of Birth ....
4. To the best of your knowledge, is there any reason for your child not taking part in this study? ....
....

5. How would you rate your child's current activity levels?

Please tick: [ ] Occasional exercise (e.g. just school lessons).
[ ] Moderate exercise (e.g. school lessons **and** some out of school activities).
[ ] Regular exercise (e.g. school lessons and frequent out of school activities.

I give consent for my child to participate in the exercise tests outlined above.

Signature of parent/guardian: ....

B2. INSTRUCTIONS FOR USING THE CERT AND RPE SCALES

CERT

"When I ask you, I want you to let me know how hard you are finding the exercise. You can do this by *pointing* to the number or the words which best describe how hard it feels, or you can *read out* the number or words to me. If the exercise feels really easy and you feel that you can keep going for a long time, then point to or tell me a low number, like 1 or 2. If it is a bit harder, but you still think you can keep going for a long time, then point to or tell me a middle number, like 4 or 5. If you feel that the exercise is getting harder, maybe you're getting hot or sweaty and breathing harder, but you can keep going, point to or tell me a higher number, like 7 or 8. If you feel it's so hard that you can't keep going much longer, point to or tell me one of the highest numbers, like 9 or 10. I want you to be *honest* - don't try to be brave and say it's easy when it's not, and it doesn't matter what any of your friends might say when it's their turn."

RPE

"During the exercise we want you to rate your perception of exertion. We want you to use this rating scale where 6 means no exertion at all and 20 means a maximal exertion. 9 is a very light exercise, like walking slowly for some minutes (for healthy people). 13 on the scale is a somewhat heavy exercise but it still feels fine and you should not have any problems to continue exercising. When you come to 17, "very hard", it is really very strenuous, you can still go on but you have to push yourself very much. 19 on the scale is an extremely strenuous exercise. For most people this is an exercise as strenuous as they have ever experienced before.

Try to appraise your feeling of exertion as honestly as possible. Don't underestimate it, but don't overestimate it either. Some people are a bit insensitive or want to be "brave" and rate too low. Don't do that but try to feel your exertion as you perceive it. Don't bother about how heavy the load is physically or what the exercise objectively might be. We are only interested in your own feeling of effort and exertion. Look at the scale and the wordings and then give us a number. You can equally well give us an even as an odd number."

[From Borg (1985; p. 25). *An Introduction to Borg's RPE-Scale.* Ithaca, New York: Mouvement.]

B3. GROUP COMPARISONS (CERT VERSUS RPE) OF STUDY 2 BIOMETRIC DATA

ALL

<u>Height</u>

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	36	135.3056	4.738	.790
Group 2	34	135.5294	4.554	.781

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.08	822	-.20	68	.841	-.20	67.98	.841

<u>Weight</u>

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	36	31.3889	4.924	.821
Group 2	34	31.9412	4.799	.823

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.05	.884	-.47	68	.636	-.48	67.93	.636

<u>Age</u>

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	36	114.9444	3.329	.555
Group 2	34	115.0588	3.923	.673

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.39	.341	-.13	68	.896	-.13	64.86	.896

<u>Physical Work Capacity</u>

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	35	99.8657	17.514	2.960
Group 2	33	104.9515	14.971	2.606

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.37	.375	-1.28	66	.204	-1.29	65.39	.202

BOYS

Height

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	16	135.4375	4.258	1.064
Group 2	12	136.0000	4.612	1.331

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.17	.757	-.33	26	.741	-.33	22.74	.744

Weight

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	16	30.5625	2.421	.605
Group 2	12	31.1667	2.552	.737

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.11	.831	-.64	26	.529	-.63	23.13	.533

Age

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	16	114.6875	3.114	.778
Group 2	12	115.4167	3.528	1.018

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.28	.640	-.58	26	.567	-.57	22.08	.575

Physical Work Capacity

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	15	109.2400	11.713	3.024
Group 2	12	111.4917	15.956	4.606

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.86	.275	-.42	25	.676	-.41	19.66	.687

GIRLS

Height

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	20	135.2000	5.197	1.162
Group 2	22	135.2727	4.610	.983

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.27	.592	-.05	40	.962	-.05	38.21	.962

Weight

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	20	32.0500	6.245	1.396
Group 2	22	32.3636	5.678	1.211

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.21	.669	-.17	40	.865	-.17	38.58	.866

Age

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	20	115.1500	3.558	.796
Group 2	22	114.8636	4.190	.893

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.39	.478	.24	40	.813	.24	39.83	.812

Physical Work Capacity

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	20	92.8350	18.067	4.040
Group 2	21	101.2143	13.350	2.913

		Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.83	.188	-1.69	39	.098	-1.68	34.93	.101

B4. REPEATED MEASURES ANOVA OF STUDY 2 DATA

HEART RATE

```
set/length=60/width=110.
get/file 'c:\spss\data\project.dat'.
manova hr25 to hr75 hr25a to hr75a by sex (1,2) scale (1,2)
/wsfactors=trials (2) Levels (3)/print=signif(averf).
```

```
67 cases accepted.
 0 cases rejected because of out-of-range factor values.
 3 cases rejected because of missing data.
 4 non-empty cells.

 1 design will be processed.
```

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	48660.13	63	772.38		
CONSTANT	9536869.34	1	9536869.3	12347.33	.000
SEX	11099.63	1	11099.63	14.37	.000
SCALE	1.93	1	1.93	.00	.960
SEX BY SCALE	693.28	1	693.28	.90	.347

Tests involving 'TRIALS' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	5824.76	63	92.46		
TRIALS	1213.99	1	1213.99	13.13	.001
SEX BY TRIALS	9.99	1	9.99	.11	.744
SCALE BY TRIALS	784.89	1	784.89	8.49	.005
SEX BY SCALE BY TRIALS	.40	1	.40	.00	.948

Tests involving 'LEVELS' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	8554.88	126	67.90		
LEVELS	187795.76	2	93897.88	1382.97	.000
SEX BY LEVELS	154.76	2	77.38	1.14	.323
SCALE BY LEVELS	111.85	2	55.92	.82	.441
SEX BY SCALE BY LEVELS	104.68	2	52.34	.77	.465

Tests involving 'TRIALS BY LEVELS' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	2787.44	126	22.12		
TRIALS BY LEVELS	95.76	2	47.88	2.16	.119
SEX BY TRIALS BY LEVELS	21.45	2	10.73	.48	.617
SCALE BY TRIALS BY LEVELS	159.55	2	79.77	3.61	.030
SEX BY SCALE BY TRIALS BY LEVELS	44.04	2	22.02	1.00	.373

PERCEIVED EFFORT

```
manova pe25 to pe75 pe25a to pe75a by sex (1,2) scale (1,2)
/wsfactors=trials (2) Levels (3)/print=signif(averf).
```

67 cases accepted.
0 cases rejected because of out-of-range factor values.
3 cases rejected because of missing data.
4 non-empty cells.
1 design will be processed.

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	945.33	63	15.01		
CONSTANT	21145.36	1	21145.36	1409.20	.000
SEX	46.45	1	46.45	3.10	.083
SCALE	4429.52	1	4429.52	295.20	.000
SEX BY SCALE	9.22	1	9.22	.61	.436

Tests involving 'TRIALS' Within-Subject Effect.

Tests of Significance for T2 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	120.75	63	1.92		
TRIALS	32.59	1	32.59	17.00	.000
SEX BY TRIALS	1.64	1	1.64	.86	.358
SCALE BY TRIALS	5.92	1	5.92	3.09	.084
SEX BY SCALE BY TRIALS	.20	1	.20	.10	.748

Tests involving 'LEVELS' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	335.96	126	2.67		
LEVELS	1169.35	2	584.68	219.28	.000
SEX BY LEVELS	.11	2	.05	.02	.980
SCALE BY LEVELS	21.51	2	10.76	4.03	.020
SEX BY SCALE BY LEVELS	13.51	2	6.76	2.53	.083

Tests involving 'TRIALS BY LEVELS' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	105.68	126	.84		
TRIALS BY LEVELS	26.88	2	13.44	16.02	.000
SEX BY TRIALS BY LEVELS	1.26	2	.63	.75	.474
SCALE BY TRIALS BY LEVELS	3.19	2	1.60	1.90	.153
SEX BY SCALE BY TRIALS BY LEVELS	2.98	2	1.49	1.78	.173

B5. REGRESSION ANALYSIS OF TRIAL 2 DATA (STUDY 2)

```
REGRESSION
  /MISSING PAIRWISE /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT hr2
  /METHOD=ENTER pe2  .
```

HEART RATE

CERT

```
Multiple R           .72610          Analysis of Variance
R Square             .52723                            DF      Sum of Squares      Mean Square
Adjusted R Square    .52308          Regression         1         46590.46484      46590.46484
Standard Error     19.14355          Residual         114         41778.22481        366.47566

                                     F =      127.13113       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B          SE B         Beta          T  Sig T

PE2            8.134888       .721482      .726105     11.275  .0000
(Constant)   127.627810      3.571155                  35.739  .0000
```

RPE

```
Multiple R           .63736          Analysis of Variance
R Square             .40623                            DF      Sum of Squares      Mean Square
Adjusted R Square    .40107          Regression         1         37400.15958      37400.15958
Standard Error     21.80263          Residual         115         54665.80623        475.35484

                                     F =       78.67840       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B          SE B         Beta          T  Sig T

PE2            4.876455       .549764      .637363      8.870  .0000
(Constant)   107.139642      6.494962                  16.496  .0000
```

POWER OUTPUT

CERT

```
Multiple R           .80035          Analysis of Variance
R Square             .64056                            DF      Sum of Squares      Mean Square
Adjusted R Square    .63740          Regression         1         43969.14421      43969.14421
Standard Error     14.71158          Residual         114         24673.09717        216.43068

                                     F =      203.15579       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B          SE B         Beta          T  Sig T

PE2            7.902728       .554450      .800347     14.253  .0000
(Constant)    20.814150      2.744388                   7.584  .0000
```

RPE

```
Multiple R           .70434          Analysis of Variance
R Square             .49610                            DF      Sum of Squares      Mean Square
Adjusted R Square    .49172          Regression         1         39353.85377      39353.85377
Standard Error     18.64382          Residual         115         39973.06930        347.59191

                                     F =      113.21856       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B          SE B         Beta          T  Sig T

PE2            5.002201       .470113      .704342     10.640  .0000
(Constant)     1.513744      5.553957                    .273  .7857
```

BOYS & GIRLS

```
GET /FILE 'c:\spss\data\pehrwl.dat'.
set/length=59/width=100.
process if (scale=1).
select if (sex=1).
regress/vars pe hr/dependent hr/method enter.
```

HEART RATE

```
process if (scale=1).
select if (sex=1).
regress/vars pe2 hr2/dependent hr2/method enter.
```

CERT: BOYS (T2)

```
Multiple R            .78735          Analysis of Variance
R Square              .61992                             DF      Sum of Squares      Mean Square
Adjusted R Square     .61275          Regression          1         26798.25632      26798.25632
Standard Error      17.60691          Residual           53         16430.18005        310.00340

                                      F =      86.44504        Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B         SE B        Beta          T  Sig T

PE2            8.788200      .945213     .787351      9.298  .0000
(Constant)   120.237162     4.773944                 25.186  .0000
```

CERT: GIRLS (T2)

```
Multiple R            .68826          Analysis of Variance
R Square              .47370                             DF      Sum of Squares      Mean Square
Adjusted R Square     .46478          Regression          1         20665.11964      20665.11964
Standard Error      19.72689          Residual           59         22959.86397        389.15024

                                      F =      53.10319        Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B         SE B        Beta          T  Sig T

PE2            7.616475     1.045186     .688258      7.287  .0000
(Constant)   133.894524     5.076439                 26.376  .0000
```

RPE: BOYS (T2)

```
Multiple R            .78599          Analysis of Variance
R Square              .61778                             DF      Sum of Squares      Mean Square
Adjusted R Square     .60868          Regression          1         20572.33600      20572.33600
Standard Error      17.40847          Residual           42         12728.30036        303.05477

                                      F =      67.88323        Signif F =  .0000

------------------ Variables in the Equation ------------------
Variable              B         SE B        Beta          T  Sig T

PE2            5.745587      .697354     .785987      8.239  .0000
(Constant)    95.172939     8.017501                 11.871  .0000
```

RPE: GIRLS (T2)

```
Multiple R            .54119          Analysis of Variance
R Square              .29289                             DF      Sum of Squares      Mean Square
Adjusted R Square     .28293          Regression          1         16826.94172      16826.94172
Standard Error      23.92044          Residual           71         40625.30486        572.18739

                                      F =      29.40810        Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B         SE B        Beta          T  Sig T

PE2            4.230773      .780165     .541189      5.423  .0000
(Constant)   116.055810     9.362866                 12.395  .0000
```

POWER OUTPUT

CERT: BOYS (T2)

Multiple R	.83619	Analysis of Variance			
R Square	.69922		DF	Sum of Squares	Mean Square
Adjusted R Square	.69354	Regression	1	27412.43557	27412.43557
Standard Error	14.91619	Residual	53	11792.10988	222.49264

F = 123.20603 Signif F = .0000

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
PE2	8.888336	.800764	.836191	11.100	.0000
(Constant)	20.143838	4.044380		4.981	.0000

CERT: GIRLS (T2)

Multiple R	.77995	Analysis of Variance			
R Square	.60832		DF	Sum of Squares	Mean Square
Adjusted R Square	.60168	Regression	1	16703.83922	16703.83922
Standard Error	13.50152	Residual	59	10755.17717	182.29114

F = 91.63275 Signif F = .0000

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
PE2	6.847676	.715349	.779948	9.572	.0000
(Constant)	21.969627	3.474428		6.323	.0000

RPE: BOYS (T2)

Multiple R	.83229	Analysis of Variance			
R Square	.69270		DF	Sum of Squares	Mean Square
Adjusted R Square	.68539	Regression	1	21725.68961	21725.68961
Standard Error	15.14843	Residual	42	9637.94675	229.47492

F = 94.67566 Signif F = .0000

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
PE2	5.904449	.606820	.832288	9.730	.0000
(Constant)	-5.052881	6.976636		-.724	.4729

RPE: GIRLS (T2)

Multiple R	.62995	Analysis of Variance			
R Square	.39684		DF	Sum of Squares	Mean Square
Adjusted R Square	.38835	Regression	1	18979.13257	18979.13257
Standard Error	20.15650	Residual	71	28846.20989	406.28465

F = 46.71388 Signif F = .0000

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
PE2	4.493195	.657404	.629954	6.835	.0000
(Constant)	5.393000	7.889598		.684	.4965

B6. CERT CONCEPTUAL MODEL[1]

```
GRAPH
 /SCATTERPLOT(BIVAR)=cert WITH hr
 /MISSING=LISTWISE
 /TEMPLATE='C:\SPSSWIN\CERTLIN.CHT'.

 REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT hr
  /METHOD=ENTER cert
```

HR:CERT

```
Variable(s) Entered on Step Number  1..     CERT


Multiple R            .99968          Analysis of Variance
R Square              .99937                            DF       Sum of Squares        Mean Square
Adjusted R Square     .99929          Regression         1          11808.10909        11808.10909
Standard Error        .96766          Residual           8              7.49091             .93636

                                      F =    12610.60194        Signif F =  .0000


------------------ Variables in the Equation ------------------

Variable              B          SE B        Beta       T       Sig T

CERT          11.963636       .106536     .999683   112.297   .0000
(Constant)   110.000000       .661037               166.405   .0000
```

Heart Rate = 110 + 11.96 (CERT)

POWER:CERT

```
Multiple R            .99996          Analysis of Variance
R Square              .99992                            DF       Sum of Squares        Mean Square
Adjusted R Square     .99991          Regression         1           8340.24545         8340.24545
Standard Error        .28604          Residual           8               .65455             .08182

                                      F =  101936.33333        Signif F =  .0000


------------------ Variables in the Equation ------------------

Variable              B          SE B        Beta          T   Sig T

CERT          10.054545       .031492     .999961   319.275   .0000
(Constant)     9.600000       .195402                49.130   .0000
```

Power Output = 9.6 + 10.1 (CERT)

[1]Calculated from Williams, J.G., Eston R.G., & Furlong, B. (1994), *Perceptual and Motor Skills*, 79, 1451-1458.

B7. CURVE ESTIMATION FOR TRIAL 2 CERT DATA (STUDY 2)

ALL

Dependent	Mth	Rsq	d.f.	F	Sigf	b0	b1	b2	b3
HR2	LIN	.527	114	127.13	.000	127.628	8.1349		
HR2	QUA	.584	113	79.23	.000	107.182	18.8268	-1.0390	
HR2	CUB	.594	112	54.59	.000	120.605	6.8071	1.7475	-.1795
HR2	POW	.553	114	140.85	.000	123.521	.2038		

Curvilinear equation: $HR = 120.61 + 6.81(CERT) + 1.75(CERT^2) - 0.18(CERT^3)$

BOYS

Dependent	Mth	Rsq	d.f.	F	Sigf	b0	b1	b2	b3
HR2	LIN	.620	53	86.45	.000	120.237	8.7882		
HR2	QUA	.675	52	54.01	.000	99.6683	19.1693	-.9769	
HR2	CUB	.688	51	37.56	.000	114.892	5.5539	2.1617	-.1992
HR2	POW	.624	53	88.00	.000	116.601	.2257		

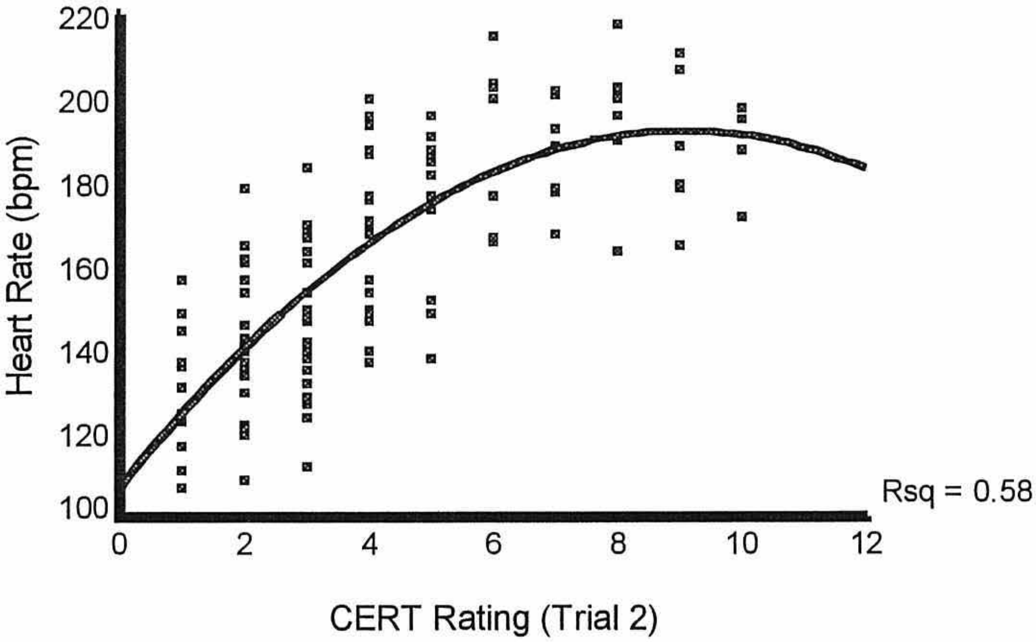
Curvilinear equation: $HR = 114.9 + 5.55(CERT) + 2.16(CERT^2) - 0.20(CERT^3)$

GIRLS

Dependent	Mth	Rsq	d.f.	F	Sigf	b0	b1	b2	b3
HR2	LIN	.474	59	53.10	.000	133.895	7.6165		
HR2	QUA	.541	58	34.22	.000	111.854	19.5638	-1.1994	
HR2	CUB	.552	57	23.36	.000	125.573	7.1132	1.7428	-.1947
HR2	POW	.521	59	64.23	.000	129.601	.1866		

Curvilinear equation: $HR = 125.6 + 7.11(CERT) + 1.74(CERT^2) - 0.19(CERT^3)$

B8. CURVILINEAR RELATIONSHIP BETWEEN HEART RATE AND CERT



C1. REPEATED MEASURES ANOVA OF STUDY 3 DATA

CERT GROUP

Heart Rate

```
FILE/A:\PRODHR.DAT.
MANOVA hrp111 hrp112 hrp113 hrp114 hrp211 hrp212 hrp213 hrp214 BY sex2(1 2)
  /WSFACTORS trials(2) levels(4)
  /METHOD UNIQUE
  /ERROR WITHIN+RESIDUAL
  /PRINT SIGNIF( AVERF )
  /NOPRINT PARAM(ESTIM) SIGNIF( MULT ).
```

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	41564.86	30	1385.50		
SEX2	357.14	1	357.14	.26	.615

Tests involving 'TRIALS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	8430.36	30	281.01		
TRIALS	1685.63	1	1685.63	6.00	.020
SEX2 BY TRIALS	282.89	1	282.89	1.01	.324

Tests involving 'LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	19949.09	90	221.66		
LEVELS	23471.83	3	7823.94	35.30	.000
SEX2 BY LEVELS	269.26	3	89.75	.40	.750

Tests involving 'TRIALS BY LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	14831.08	90	164.79		
TRIALS BY LEVELS	231.84	3	77.28	.47	.705
SEX2 BY TRIALS BY LEVELS	359.51	3	119.84	.73	.538

RPE GROUP

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	53484.85	32	1671.40		
SEX2	808.18	1	808.18	.48	.492

Tests involving 'TRIALS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	14481.98	32	452.56		
TRIALS	1295.75	1	1295.75	2.86	.100
SEX2 BY TRIALS	624.77	1	624.77	1.38	.249

Tests involving 'LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	18874.78	96	196.61		
LEVELS	49328.50	3	16442.83	83.63	.000
SEX2 BY LEVELS	288.10	3	96.03	.49	.691

Tests involving 'TRIALS BY LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	11227.32	96	116.95		
TRIALS BY LEVELS	273.97	3	91.32	.78	.507
SEX2 BY TRIALS BY LEVELS	140.37	3	46.79	.40	.753

CERT GROUP

POWER OUTPUT

```
MANOVA  wlp111  wlp112  wlp113  wlp114  wlp211  wlp212  wlp213  wlp214  BY
sex2(1 2)
 /WSFACTORS trials(2) levels(4)
 /METHOD UNIQUE
 /ERROR WITHIN+RESIDUAL
 /PRINT SIGNIF( MULT AVERF )
 /NOPRINT PARAM(ESTIM) .
```

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	34927.94	30	1164.26		
SEX2	18215.50	1	18215.50	15.65	.000

Tests involving 'TRIALS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	15940.50	30	531.35		
TRIALS	2611.61	1	2611.61	4.92	.034
SEX2 BY TRIALS	23.99	1	23.99	.05	.833

Tests involving 'LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	29985.90	90	333.18		
LEVELS	79187.50	3	26395.83	79.22	.000
SEX2 BY LEVELS	2342.13	3	780.71	2.34	.078

Tests involving 'TRIALS BY LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	25099.36	90	278.88		
TRIALS BY LEVELS	558.62	3	186.21	.67	.574
SEX2 BY TRIALS BY LE VELS	1204.31	3	401.44	1.44	.237

RPE GROUP

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	48721.44	32	1522.54		
SEX2	4522.06	1	4522.06	2.97	.094

Tests involving 'TRIALS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	18031.39	32	563.48		
TRIALS	3806.08	1	3806.08	6.75	.014
SEX2 BY TRIALS	198.14	1	198.14	.35	.557

Tests involving 'LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	31611.75	96	329.29		
LEVELS	99658.63	3	33219.54	100.88	.000
SEX2 BY LEVELS	1055.98	3	351.99	1.07	.366

Tests involving 'TRIALS BY LEVELS' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	20942.44	96	218.15		
TRIALS BY LEVELS	2713.56	3	904.52	4.15	.008
SEX2 BY TRIALS BY LEVELS	125.74	3	41.91	.19	.902

C2. POST-HOC (TUKEY) ANALYSIS: WORKED EXAMPLE ON STUDY 3 DATA

$$T_{0.05} = q_{(n,\,V)} \times \sqrt{\frac{MSw}{N}}$$

where MS_w = within-subjects variance and N = number in each group or the number of scores from which each value is calculated.

q is obtained from tables (see Cohen & Holliday, 1979; pp. 208-211), using the n (number of means to be compared) and V (degrees of freedom corresponding to within-subjects variance) values.

CERT Scale

Heart Rate

Significant Main Effect of Levels (3, 5, 7, & 9), $F(3, 90) = 35.3$; $p < .0001$

Level	Mean	diff.	sig (.05)	sig (.01)
3	162.2			
		12.5		*
5	174.7			
		7.6	*	
7	182.3			
		6.0	ns	
9	188.3			

MSw = 221.66
V = 90
n = 4
q (.05) = 3.715 (by interpolation)
q (.01) = 4.55

$T = 3.715 \times \sqrt{(221.66 / 67)} = 6.76$
$T = 4.550 \times \sqrt{(221.66 / 67)} = 8.28$

Power Output

Significant Main Effect of Levels (3, 5, 7, & 9), $F(3, 90) = 38.3$; $p < .0001$

Level	Mean	diff.	sig (.05)	sig (.01)
3	53.0			
		16.8		*
5	70.2			
		10.7		*
7	80.9			
		21.7		*
9	102.6			

MSw = 375.87
V = 90
n = 4
q (.05) = 3.715 (by interpolation)
q (.01) = 4.55

$T = 3.715 \times \sqrt{(375.87 / 67)} = 8.80$
$T = 4.550 \times \sqrt{(375.87 / 67)} = 10.7$

RPE Scale

Heart Rate

Significant Main Effect of Levels (8, 12, 15, & 18, $F(3, 96) = 83.6$; $p < .0001$

Level	Mean	diff.	sig (.05)	sig (.01)
8	145.6			
		8.7		*
12	164.3			
		12.0		*
15	176.3			
		5.6	ns.	
18	188.3			

MSw = 196.61
V = 96
n = 4

q (.05) = 3.735 (by interpolation)
q (.01) = 4.59

$T = 3.735 \times \sqrt{(196.61 / 68)} = 6.35$
$T = 4.590 \times \sqrt{(196.61 / 68)} = 7.80$

Power Output

Significant Main Effect of Levels (8, 12, 15, & 18), $F(3, 96) = 100.9$; $p < .0001$

Level	Mean	diff.	sig (.05)	sig (.01)
8	36.8			
		14.3		*
12	51.1			
		15.8		*
15	67.3			
		16.8		*
9	84.1			

MSw = 329.29
V = 96
n = 4
q (.05) = 3.735 (by interpolation)
q (.01) = 4.55

$T = 3.735 \times \sqrt{[(329.29/2)(1/67 + 1/68)]} = 8.25$ (Uneven groups)
$T = 4.550 \times \sqrt{[(329.29/2)(1/67 + 1/68)]} = 10.14$

D1. REPEATED MEASURES ANOVA OF STUDY 4 DATA

HEART RATE

```
 66 cases accepted.
  0 cases rejected because of out-of-range factor values.
198 cases rejected because of missing data.
  4 non-empty cells.

        1 design will be processed.
```

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	92817.07	62	1497.05		
GROUP	4603.64	1	4603.64	3.08	.084
SEX	7121.88	1	7121.88	4.76	.033
GROUP BY SEX	468.75	1	468.75	.31	.578

Tests involving 'LEVELS' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	52093.24	186	280.07		
LEVELS	65305.10	3	21768.37	77.72	.000
GROUP BY LEVELS	3253.24	3	1084.41	3.87	.010
SEX BY LEVELS	549.75	3	183.25	.65	.581
GROUP BY SEX BY LEVELS	260.65	3	86.88	.31	.818

Tests involving 'TIME' Within-Subject Effect.

Tests of Significance for T5 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1384.26	62	22.33		
TIME	3884.62	1	3884.62	173.99	.000
GROUP BY TIME	306.94	1	306.94	13.75	.000
SEX BY TIME	70.16	1	70.16	3.14	.081
GROUP BY SEX BY TIME	30.60	1	30.60	1.37	.246

Tests involving 'LEVELS BY TIME' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	2705.75	186	14.55		
LEVELS BY TIME	559.10	3	186.37	12.81	.000
GROUP BY LEVELS BY TIME	84.53	3	28.18	1.94	.125
SEX BY LEVELS BY TIME	44.31	3	14.77	1.02	.387
GROUP BY SEX BY LEVELS BY TIME	64.05	3	21.35	1.47	.225

POWER OUTPUT

```
 66 cases accepted.
  0 cases rejected because of out-of-range factor values.
198 cases rejected because of missing data.
  4 non-empty cells.

  1 design will be processed.
```

Tests of Between-Subjects Effects.

Tests of Significance for T1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	33653.45	62	542.80		
GROUP	951.52	1	951.52	1.75	.190
SEX	1387.31	1	1387.31	2.56	.115
GROUP BY SEX	4444.59	1	4444.59	8.19	.006

Tests involving 'LEVELS' Within-Subject Effect.

AVERAGED Tests of Significance for T1L using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	42345.29	186	227.66		
LEVELS	107385.37	3	35795.12	157.23	.000
GROUP BY LEVELS	2631.85	3	877.28	3.85	.010
SEX BY LEVELS	855.19	3	285.06	1.25	.292
GROUP BY SEX BY LEVELS	146.63	3	48.88	.21	.886

D2. COMPARISON OF PEARSON CORRELATIONS USING FISHER'S Z-TRANSFORMATION[1]

Example Calculation:

Comparison of correlation coefficients between heights and 100 m sprint times of 14-year-old and 18-year-old boys.

14-year-olds, n = 50, r = -0.75
18-year-olds, n = 28, r = -0.63

To test whether these two correlations are significantly different, they are converted to Z-scores: $Zr = 0.5 \times \log_e [(1 + r)/(1 - r)]$

i.e. Zr = 0.973 for 14-year-olds
Zr = 0.741 for 18-year-olds

It is the ratio of the difference between these two Z scores to the standard deviation of the difference which is then tested for significance (in the manner of a *t*-test).

i.e. $t = (Zr_1 - Zr_2) / Szr_1 - zr_2$ where $Szr_1 - zr_2 = (1 / n_1 - 3) + (1 / n_2 - 3)$

$Szr_1 - zr_2 = (1/47) + (1/25) = 0.061276 = 0.248$

Therefore, t = **(0.973 - 0.741) / 0.248 = 0.935**

The critical value of t with 72 degrees of freedom ($n_1 + n_2 - 6$) at the 0.05 level of significance = 1.993

Since the above value is less that 1.993, it can be concluded that the correlation between height and sprinting time is similar in the 14-year-old and 18-year-old boys.

N.B. An example from Study 2 data (correlations between heart rates and CERT ratings at two different time periods) follows.

---

[1]Calculation adapted from an example described in Morehouse & Stull (1975; pp. 201-204), *Statistical Principles and Procedures with Applications for Physical Education*. Philadelphia: Lea & Febiger.

EXAMPLE DATA FROM STUDY 4

Heart Rate Data

Comparison	n	r	Zr	σZr^2	$SZr_1 - Zr_2$	t	df	t_{crit}	signif.
All @2min									
Continuous	132	0.41	0.435	1/129	0.125	2.472	258	1.650	<.05
Discontinuous	132	0.63	0.744	1/129					
Boys @2min									
Continuous	84	0.47	0.510	1/81	0.159	1.67	158	1.654	<.05
Discontinuous	80	0.65	0.775	1/77					
Girls @2min									
Continuous	48	0.36	0.375	1/45	0.206	1.699	94	1.662	<.05
Discontinuous	52	0.62	0.725	1/49					
All @3min									
Continuous	132	0.66	0.795	1/129	0.125	2.370	258	1.650	<.05
Discontinuous	132	0.46	0.500	1/129					
Boys @3min									
Continuous	84	0.50	0.550	1/81	0.159	1.874	158	1.654	<.05
Discontinuous	80	0.69	0.848	1/77					
Girls @3min									
Continuous	48	0.45	0.485	1/45	0.206	1.408	94	1.662	NS
Discontinuous	52	0.65	0.775	1/49					

(cont.)

Power Output Data[2]

Comparison	n	r	Zr	σZr^2	$SZr_1 - Zr_2$	t	df	t_{crit}	signif.
All									
Continuous	132	0.74	0.950	1/129	0.125	0.560	258	1.650	NS
Discontinuous	132	0.77	1.020	1/129					
Boys									
Continuous	84	0.74	0.950	1/81	0.159	1.922	158	1.654	<.05
Discontinuous	80	0.85	1.256	1/77					
Girls									
Continuous	48	0.76	0.995	1/45	0.206	0.799	94	1.662	NS
Discontinuous	52	0.68	0.830	1/49					

[2]Based on all values recorded after 2 min

E1. TEST-RETEST RELIABILITY ANALYSIS USING BLAND & ALTMAN'S 95% LIMITS OF AGREEMENT[1]

STUDY 1 HEART RATE DATA

	BIAS (mean diff)	*S.D.*	*95% LIMITS (+/- b/min)*
All			
CERT 5	1.9	9.5	18.6
CERT 7	2.5	11.6	22.8
CERT 9	2.9	11.5	22.4
OVERALL	**2.4**	**10.7**	**20.9**
Boys			
CERT 5	5.0	12.3	24.1
CERT 7	5.4	14.2	27.8
CERT 9	4.4	15.7	30.7
OVERALL	**4.9**	**13.5**	**26.4**
Girls			
CERT 5	-1.3	4.4	8.6
CERT 7	-0.4	8.3	16.3
CERT 9	1.4	5.8	11.3
OVERALL	**-0.1**	**6.2**	**12.2**

STUDY 1 POWER OUTPUT DATA

	BIAS (mean diff)	*S.D.*	*95% LIMITS (+/- watts)*
All			
CERT 5	0.6	6.3	12.3
CERT 7	-0.3	6.2	12.1
CERT 9	0.6	12.9	25.3
OVERALL	**0.3**	**8.8**	**17.3**
Boys			
CERT 5	3.8	5.2	10.2
CERT 7	2.5	5.4	10.5
CERT 9	6.9	12.2	24.0
OVERALL	**4.4**	**8.1**	**15.9**
Girls			
CERT 5	-2.5	6.0	11.7
CERT 7	-3.1	2.1	4.1
CERT 9	-5.6	10.8	21.3
OVERALL	**-3.8**	**7.7**	**15.1**

STUDY 2 EFFORT RATINGS DATA: CERT SCALE

INTENSITY (watts)	*BIAS (mean diff)*	*S.D.*	*95% LIMITS (+/- rating)*
All			
25	0.7	1.2	2.3
50	0.5	1.3	2.5
75	-0.2	1.3	2.6
100	0.0	1.5	2.9
OVERALL	**0.3**	**1.3**	**2.6**
Boys			
25	0.5	1.0	1.9
50	0.7	1.3	2.5
75	0.3	1.2	2.3
100	0.0	1.5	2.9
OVERALL	**0.4**	**1.2**	**2.4**
Girls			
25	0.8	1.3	2.5
50	0.4	1.3	2.5
75	-0.6	1.4	2.7
OVERALL	**0.2**	**1.4**	**2.7**

STUDY 2 EFFORT RATINGS DATA: RPE SCALE

INTENSITY (watts)	*BIAS (mean diff)*	*S.D.*	*95% LIMITS (+/- rating)*
All			
25	1.4	1.6	3.1
50	1.1	1.6	3.2
75	-0.1	2.1	4.1
100	0.3	2.1	4.0
OVERALL	**0.8**	**1.9**	3.8
Boys			
25	1.5	1.6	3.2
50	1.6	1.3	2.6
75	-0.3	2.2	4.2
100	0.3	2.6	5.1
OVERALL	**0.8**	**2.0**	**3.9**
Girls			
25	1.4	1.6	3.2
50	0.9	1.8	3.5
75	0.0	2.1	4.2
100	0.3	1.7	3.3
OVERALL	0.7	1.9	3.7

STUDY 3 HEART RATE DATA (CERT SCALE)

	BIAS (mean diff)	*S.D.*	*95% LIMITS (+/- b/min)*
All			
CERT 3	6.7	26.0	51.0
CERT 5	4.0	20.8	40.7
CERT 7	8.0	18.6	36.5
CERT 9	3.1	9.4	18.3
OVERALL	**5.4**	**19.5**	**38.2**
Boys			
CERT 3	8.0	22.9	45.0
CERT 5	2.4	22.3	43.8
CERT 7	1.9	16.8	32.9
CERT 9	0.0	7.0	13.7
OVERALL	**3.1**	**18.2**	**35.6**
Girls			
CERT 3	5.6	28.8	56.5
CERT 5	5.3	20.0	39.2
CERT 7	12.7	19.1	37.4
CERT 9	5.6	10.4	20.3
OVERALL	**7.3**	**20.4**	**40.0**

STUDY 3 POWER OUTPUT DATA (CERT SCALE)

	BIAS (mean diff)	*S.D.*	*95% LIMITS (+/- watts)*
All			
CERT 3	7.0	30.5	60.8
CERT 5	5.2	26.8	52.5
CERT 7	9.8	22.0	43.1
CERT 9	2.4	24.5	48.1
OVERALL	**7.1**	**25.2**	**49.4**
Boys			
CERT 3	9.6	31.2	61.2
CERT 5	8.4	34.5	67.5
CERT 7	2.7	22.2	43.5
CERT 9	-1.4	18.8	36.9
OVERALL	**5.8**	**26.5**	**51.8**
Girls			
CERT 3	4.9	32.5	63.6
CERT 5	2.7	19.6	38.4
CERT 7	15.3	20.7	40.7
CERT 9	5.3	28.4	55.6
OVERALL	**8.1**	**24.3**	**47.6**

STUDY 3 HEART RATE DATA (RPE SCALE)

	BIAS (mean diff)	*S.D.*	*95% LIMITS (+/- b/min)*
All			
RPE 8	6.8	21.9	42.9
RPE 12	6.9	24.3	47.6
RPE 15	6.6	18.2	35.7
RPE 18	1.6	14.3	28.1
OVERALL	**5.5**	**19.9**	**39.1**
Boys			
RPE 8	1.9	16.2	31.7
RPE 12	0.1	23.7	46.5
RPE 15	4.8	14.0	27.4
RPE 18	-1.3	9.8	19.2
OVERALL	**1.4**	**16.3**	**32.0**
Girls			
RPE 8	9.5	24.4	47.8
RPE 12	10.7	24.3	47.6
RPE 15	7.6	20.4	40.0
RPE 18	3.2	16.3	31.9
OVERALL	**7.7**	**21.4**	**42.0**

STUDY 3 POWER OUTPUT DATA (RPE SCALE)

	BIAS (mean diff)	*S.D.*	*95% LIMITS (+/- watts)*
All			
RPE 8	8.7	21.8	42.6
RPE 12	16.9	26.4	51.8
RPE 15	9.9	26.2	51.4
RPE 18	-2.1	22.7	44.5
OVERALL	**8.4**	**25.1**	**49.1**
Boys			
RPE 8	4.3	18.7	36.6
RPE 12	13.3	22.7	44.4
RPE 15	10.1	24.3	47.7
RPE 18	-3.8	22.2	43.5
OVERALL	**6.0**	**22.3**	**43.7**
Girls			
RPE 8	11.2	23.3	45.7
RPE 12	18.9	28.6	56.1
RPE 15	9.7	27.7	54.3
RPE 18	-1.1	23.4	45.9
OVERALL	**9.6**	**26.5**	**51.9**

[1]This form of analysis considers the level of agreement of repeated measures amongst *individuals* within a sample. It is unique amongst other forms of reliability analysis in that it does not rely upon a correlation coefficient to quantify the degree of reliability (agreement). Instead, it calculates the extent of the agreement (or disagreement) for 95% of the sample (approximately +/- 2 S.D. from the mean difference) in units of the dependent variable and then allows the investigator to use his/her judgement to decide if the repeated measurements 'adequately' agree or not. Therefore, *context* is important in the decision making process.

The above data has been tabulated for economy; it could have been presented as a series of 113 so-called Bland and Altman plots, which graphically display the bias and 95% limits of agreement. However, the final decision about the reliability of a measurement (or 'tool') is still based on the *magnitude* of these two parameters. As an example from the above statistics, the interpreter must decide whether the limits of agreement for power output production using CERT in Study 3 is acceptable. Overall, the bias was quite small - only 7.1 watts - yet individuals in the sample had biases of 49.4 watts or greater, making it very difficult on first impression to conclude generally that the tool in question is reliable. But, the investigator must decide in the context of the usefulness of this particular tool whether such a trial-to-trial disagreement is, in fact, not too bad after all.