Original Article

Reliability of Effort Production Using the Children's CALER and BABE Perceived Exertion Scales

Gaynor Parfitt[1], Pam Shepherd[2], Roger G. Eston[1]

[1]*School of Sport and Health Sciences, University of Exeter, Exeter, UK*

[2]*Edexcel, High Holborn, London, UK*

The purpose of this study was to evaluate the validity and reliability of the Cart and Load Effort Rating (CALER) and Bug and Bag Effort (BABE) rating scales for intra- and intermodal regulation of effort production in a triple repeated, randomized, intermittent production paradigm. Thirty boys and girls (age, 9.8 ± 1.5 years; height, 1.39 ± 0.08 m; weight, 31.6 ± 5.80 kg) volunteered to participate in the study and were split into two groups (a CALER group and a BABE group). Each group completed six effort production trials using their allocated scales; three stepping and three cycling. Effort production levels were set at 3, 5 and 8 for each scale. The children had 2 minutes at each intensity to request an increase or decrease in load (either in the backpack or on the bike) to achieve the effort production level set before continuing at that intensity for an additional minute. Heart rate was recorded at the end of each 3-minute period. Analysis of variance was used to examine the validity of the scales, and intraclass correlation, limits of agreement, and coefficient of variation were used to examine the reliability of the scales. The results indicated that both scales were valid and reliable and that they could be used interchangeably—the CALER scale in a stepping mode and the BABE scale in a cycling mode.

Keywords: children, exercise, perceived exertion, validity

Introduction

There are numerous ways of assessing physical activity and exercise intensity in children. These include heart rate (HR), accelerometry, observation, pedometry, recall, and the rating of perceived exertion. Application of the ratings of perceived exertion to self-regulate the intensity of exercise in school physical education lessons was suggested over 20 years ago (Eston 1984). Research on perceived exertion has provided compelling arguments that adult-derived methods and applications of the rating of perceived exertion (RPE) notion is not appropriate for use with young children. Investigations on effort perception in children should consider the ability of children to understand and interpret the scale used. This has led to the modification of existing scales or the development of new ones (Eston & Parfitt 2006; Leung et al. 2002; Utter et al. 2002; Eston & Lamb 2000; Eston et al. 2000; Robertson et al. 2000; Eston et al. 1994; Williams et al. 1994).

In the last 15 years, there have been important advances in the study of effort perception in children. During this period, there has been greater focus on the creation of more developmentally appropriate scales which use meaningful terminology and symbols. These have been reviewed and described previously by Eston & Lamb (2000). The Children's Effort Rating Table (CERT) introduced in 1994 (Eston et al. 1994; Williams et al.

Corresponding Author
Roger G. Eston, School of Sport and Health Sciences, St Luke's Campus, Heavitree Road, Exeter EX1 2LU, UK.
Tel: (44) 1392 264720
Fax: (44) 1392 264706
E-mail: R.G.Eston@exeter.ac.uk

1994), and upon which the wording and structure of the most popular children's perceived exertion scales are based, is a 1–10 verbally anchored scale using words generated by children. Published examples include the 0–10 OMNI Scale that depicts a child riding a bicycle at various stages of exertion on a linear slope (Robertson et al. 2000), or walking or running up a linear slope (Utter et al. 2002), and the 1–10 Cart and Load Effort Rating (CALER) scale (Eston et al. 2000), which depicts a child on a bicycle at various stages of exertion towing a cart with an increasing load on a horizontal surface (Figure 1). We have also explored the efficacy of a stepping scale (Bug and Bag Effort [BABE] rating scale). This scale depicts a cartoon character based on a popular film, at various stages of exertion stepping up and down onto a bench whilst carrying a backpack that is progressively loaded with rocks (Eston et al. 2001) (Figure 2). Common features of all of these scales are fewer possible responses, a range of numbers (e.g. 1–10) which is more familiar to children than the Borg 6–20 Category Scale (Borg 1998) and verbal expressions chosen by children as descriptors of exercise effort.

At times in a physical education setting, it may be desirable to encourage children to self-regulate exercise at an appropriate and desirable intensity for a given length of time. Given the observation that young children's activity tends to be transitory and intermittent in nature, it is important to confirm the reliability of a scale to assess this type of activity. Limited research exists which has assessed the validity of using developmentally appropriate perceived exertion scales to self-regulate exercise intensity in young children. A few studies have assessed the validity of children's perceived exertion scales against objective markers of physiologic strain (e.g. HR, oxygen uptake), using continuous, incremental, perceptual estimation paradigms in which the child passively rates the intensity of the exercise. This paradigm has been employed to assess the validity of mode-specific OMNI Scales for cycling (Robertson et al. 2000) and treadmill walking and running (Utter et al. 2002). A criticism of such studies is the inevitable realization by the child that the exercise is getting progressively harder.

We believe that studies which have sought to emulate the intermittent nature of exercise behavior in children,

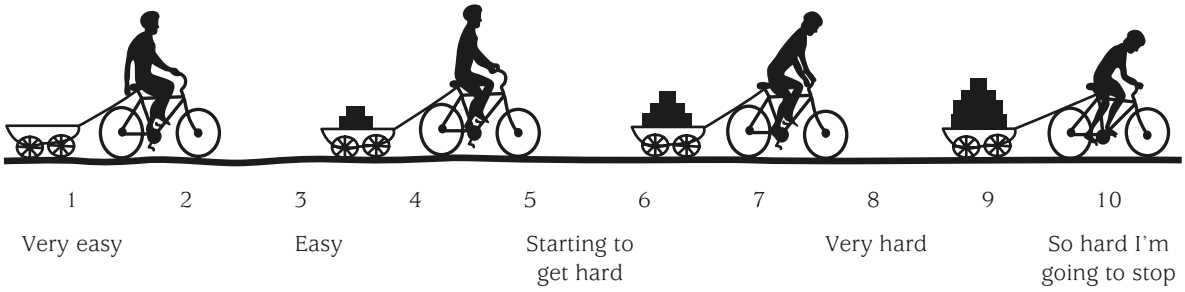

Fig. 1 Cart and Load Effort Rating (CALER) scale (Eston et al. 2000).

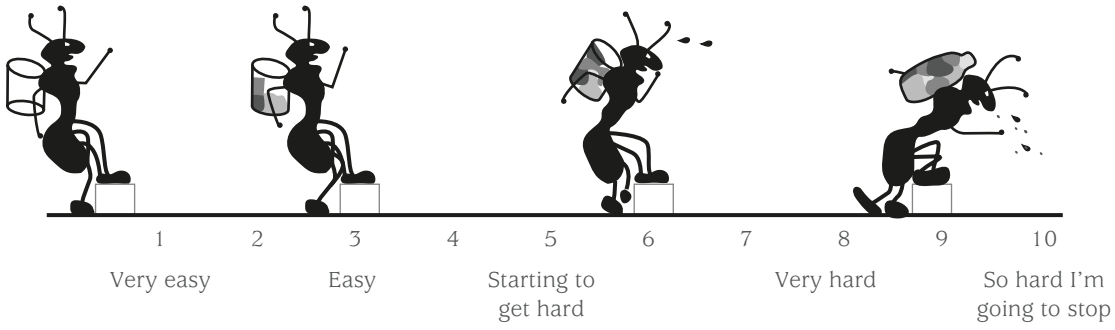

Fig. 2 Bug and Bag Effort (BABE) rating scale (Eston et al. 2001).

and which have sought to assess the validity of perceived exertion within a production paradigm setting, are more appropriate for the validation of child-specific perceived exertion scales. Studies by Lamb et al. (1997), Lamb (1996), Eston et al. (2001, 2000, 1994), and Robertson et al. (2002) are good examples, although only three of these studies have randomized or alternated the order of RPE production levels (Robertson et al. 2002; Eston et al. 2001; Lamb et al. 1997).

For a child to perceive effort accurately and then to reliably self-regulate exercise intensity at a given RPE, it is logical to assume that learning must occur. Implicit in the process of learning is repetition and practice of the skill and the cognitive ability of the child. However, few investigations on perceived exertion in children have incorporated all of these issues into their design.

Furthermore, in the development of child-specific effort rating scales, there appear to be no studies which have questioned the intermodal validity of scales. Researchers have tended to use mode-specific pictorial scales to assess their validity and reliability for cycling (Eston et al. 2000; Robertson et al. 2000), or treadmill walking or running (Utter et al. 2002). Questions have remained unanswered regarding the appropriateness of using such scales with other exercise modes, such as stepping. In this regard, it is unknown whether the BABE scale, designed specifically to assess perceived exertion by children during stepping protocols, could be used for cycling tasks. The importance (and association) that children may make between the pictorial representation depicted on the scales and the exercise task to be completed could be a factor that hinders or improves a child's reliability of effort production across trials.

The purpose of this study was to evaluate the validity and reliability of the CALER and BABE scales for intra- and intermodal regulation of effort production in a triple repeated, randomized, intermittent, production paradigm using cycling and stepping protocols.

Methods

Subjects

Thirty clinically normal boys and girls, aged 7–11 years from a primary school in Wiltshire, England, volunteered for the study. Descriptive statistics for age, height and weight were (mean ± standard deviation, SD) 9.8 ± 1.5 years, 1.39 ± 0.08 m and 31.6 ± 5.80 kg, respectively. Parental informed consent was given for each child. The study was approved by the Swindon Research Ethics Committee, Wiltshire Health Authority, Devizes, UK.

General procedures

One week prior to testing, the same experimenter (PS) habituated the children to the exercise equipment and the testing protocols to be used in the study. The children were also introduced to the CALER and BABE scales (Figures 1 & 2). The children were randomly allocated into two groups: group 1 (CALER) and group 2 (BABE). 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 scales to keep. Data collection for both groups took place during the normal school day on the school premises. The testing environment was kept constant at a comfortable 22°C by use of a fan-assisted air cooling system.

Discontinuous effort production protocols

Each group performed a discontinuous effort production protocol on six occasions, 1 week apart. For both the CALER and BABE groups, the first three trials were on a cycle ergometer. The second three trials consisted of a stepping protocol with and without a loaded backpack. All children were tested individually. Immediately prior to each effort production trial, the child was re-familiarized with either the CALER scale (group 1) or the BABE scale (group 2) and given standard instructions concerning its use and the purpose of the test.

Cycle ergometry

On each of the three production trials, the child was instructed to regulate exercise intensity on a mechanically braked cycle ergometer (Monark 824E; Monark AB, Varberg, Sweden) to match a range of three randomly presented effort production levels of 3, 5 and 8. Each production trial commenced with a 3-minute warm-up at 50 W followed by a 2.5-minute rest. Each child then instructed the experimenter to adjust the cycling resistance (to add or subtract weights in multiples of 0.1, 0.5 or 1.0 kg units), in accordance with the specified perceived levels. A metronome (DM33; Seiko, Taiwan) was

used to ensure a constant pedaling rate of 50 rpm. A "shield" was placed to hide the weights being applied to the basket by the experimenter throughout the testing. The child was allowed 2 minutes to settle on the appropriate resistance, before cycling for a further 1 minute at the designated intensity. HR (CE0537; Polar Fitwatch, Oy, Finland) and load were recorded at the end of the 3rd minute. Exercise bouts were interspersed with a 2.5-minute rest period. Following completion of the protocol, each child performed a 3-minute cool-down, cycling at 25–50 rpm against no resistance.

Discontinuous stepping protocol

After a 3-minute warm-up, each child performed an intermittent effort production protocol. This comprised stepping on and off a 0.30 m gymnasium bench. Cadence was regulated at 25 steps per minute by a metronome (DM33; Seiko). Exercise intensity was controlled by increasing the load contained in the backpack fitted to each child. The loads were calculated according to 0%, 5%, 10%, 15%, 20% and 25% body mass (Williams et al. 1994). In the first 2 minutes, exercise intensity was adjusted by instruction from the child to the investigator to add or remove weight from the backpack until he/she was confident that the randomly assigned effort rating level of 3, 5 or 8 was attained. The child then continued to step for a further 1 minute. HR (CE0537; Polar Fitwatch) and load were recorded at the end of the 3rd minute. Exercise bouts at RPEs 3, 5 and 8 were interspersed with a 2.5-minute rest period.

Scale preference

One week after the final data collection, all children were issued with a questionnaire to determine if there was a preference for a particular scale. The purpose of this simple questionnaire was to collect quantitative data to confirm if effort rating scale preference existed. The questionnaire was issued under examination conditions to the children in the classroom by their form tutor who explained the need to answer the question truthfully. Children were asked, "If you were to complete the exercises again, which scale would you prefer to use—the CALER or BABE scale?" Choice was indicated by placing a tick next to a photocopy-reduced version of the appropriate effort rating scale.

Data analysis

Data were analyzed using three-factor analysis of variance (ANOVA) [Group (2) × Mode (2) × Trial (3)] on HR, with repeated measures on Mode and Trial, at each of the three effort rating levels (3, 5, 8). In addition, a three-factor ANOVA on trial 3 data [Group (2) × Mode (2) × RPE Level (3)] was conducted on HR, with repeated measures on Mode and RPE Level. *Post hoc* comparisons were analyzed where appropriate using Tukey's test in order to identify any significant interactions or main effects. Intraclass correlation coefficients (ICCs) and the 95% limits of agreement (LoA) were calculated to give a quantitative indicator of the overall test–retest reliability (Trial 1 *vs.* Trial 2 *vs.* Trial 3) of effort production between trials using the CALER and BABE scales. The LoA calculated the mean difference (bias) and the 95% limits of agreement (±2 SD of the bias) between repeated trials (Bland & Altman 1986).

Following indications from a preliminary study which observed that children aged 7–10 years seemed to prefer using the BABE scale, a χ^2 analysis was performed on questionnaire data in order to explore the children's scale preference. The independent variables for the χ^2 analysis were "Scale Used" and "Scale Preference".

Results

A series of *t* tests confirmed that there were no significant differences between the biometric data of the two groups (CALER and BABE). Group mean (±SD) HRs at each of the three effort rating levels (3, 5, 8) for the stepping and cycling modes are shown in Table 1.

Analysis of HR differences between the three effort rating levels

A Group (2) × Mode (2) × RPE Level (3) ANOVA on trial 3 data revealed a highly significant main effect of Level on HR ($F_{2,56} = 1198.69$, $p < 0.001$). Tukey's *post hoc* analyses revealed that differences between HR at RPE 3, 5 and 8 (155 ± 16, 167 ± 16, 177 ± 15 beats min^{-1}, respectively) were significant ($p < 0.001$). ANOVA also revealed a significant main effect for Mode ($F_{1,28} = 5.50$, $p < 0.05$), with higher overall mean HR produced in the cycling protocol (169 ± 19 cf. 164 ± 17). There was also a significant interaction of Mode × Level on HR

Table 1. Heart rate (HR) by perceived effort level and Trial (Cart and Load Effort Rating [CALER] or Bug and Bag Effort [BABE] rating)*

Effort level	Trial	Cycling protocol		Stepping protocol	
		CALER HR (beats min^{-1})	BABE HR (beats min^{-1})	CALER HR (beats min^{-1})	BABE HR (beats min^{-1})
3	1	157.3 (18.8)	152.3 (19.8)	154.4 (12.3)	155.8 (13.8)
	2	152.3 (16.6)	157.8 (18.7)	153.1 (14.8)	154.4 (13.2)
	3	156.7 (13.9)	158.9 (18.7)	150.8 (15.4)	158.0 (16.3)
5	1	171.1 (16.6)	171.6 (17.3)	163.6 (13.3)	164.3 (15.6)
	2	164.3 (15.3)	170.1 (19.7)	162.8 (14.8)	163.1 (13.4)
	3	167.9 (13.5)	170.8 (17.0)	161.9 (17.7)	167.0 (18.5)
8	1	181.3 (13.4)	183.5 (17.6)	173.5 (13.5)	174.2 (15.5)
	2	178.3 (12.6)	182.1 (17.6)	173.0 (16.4)	172.7 (11.6)
	3	180.7 (12.2)	182.1 (15.2)	171.1 (16.9)	177.3 (17.1)
Overall	1	169.9 (18.9)	169.1 (22.1)	163.8 (15.0)	164.8 (16.5)
	2	165.0 (18.1)	170.0 (20.8)	163.0 (17.1)	163.4 (14.6)
	3	168.4 (16.3)	170.6 (19.2)	161.3 (18.3)	167.4 (18.7)

*Values are expressed as mean (standard deviation).

($F_{2,56} = 10.99$, $p < 0.01$). *Post hoc* analysis revealed significantly lower HR for the stepping protocol ($p < 0.01$) at RPEs 5 and 8. There were no differences at RPE 3. The increase in HR was therefore moderated by the mode of exercise, being greater for cycling compared to stepping. There were no significant differences in HRs between the CALER and BABE groups.

Analysis of HR differences between trials

A Group (2) × Mode (2) × Trials (3) ANOVA on HR was applied to each of the three RPE levels (3, 5, 8). This analysis was chosen in order to determine consistencies of exercise intensities within trials.

There was no significant difference across trials between HR produced at RPE 3 for stepping and cycling modes. There were also no significant main effects or interactions at this effort rating level. At RPE 5, ANOVA revealed a significant main effect for Mode ($F_{1,28} = 5.61$, $p < 0.05$). The mean HR produced for cycling was significantly higher (169 ± 16 cf. 164 ± 15 beats min^{-1}). The same was true for RPE 8 ($F_{1,28} = 13.01$, $p < 0.01$). The mean HR produced for cycling was again significantly higher (181 ± 15 cf. 174 ± 15 beats min^{-1}). There were no other main effects or interactions.

Reliability of effort production

Table 2 shows the ICCs for HR at the three RPE levels (3, 5, 8) for the CALER and BABE groups across trials for cycling and stepping. For cycling, the ICCs between Trial 2 and Trial 3 tended to be higher than between Trial 1 and Trial 2 across the three RPE levels for both groups. The overall ICCs for cycling were highest in the BABE group. For the stepping task, the ICCs show highly reliable HR production at all RPE levels for both the CALER and BABE groups. Table 3 shows the LoA analyses. LoAs improved across trials in the cycling mode for both scales, but was variable in the stepping mode.

Scale preference

Of the 30 children who participated in the study, 29 questionnaires were returned. The questionnaire results revealed that 3/15 (20%) children who had previously used the BABE scale chose the CALER; 9/14 (64%) children who had previously used the CALER chose the BABE scale; 12/15 (80%) children who had previously used the BABE scale still preferred to use the BABE scale; and 5/14 (36%) children who had previously used the CALER scale still preferred to use the CALER scale.

The χ^2 analyses were conducted on the data and revealed values of 4.17 and 3.84 with 1degree of freedom ($p < 0.05$). The independent variables for the χ^2 analyses were Scale Used and Scale Preference. The results indicated a significant difference ($p < 0.05$) between the CALER and BABE groups with regards to preferred choice of effort rating scale. The BABE group showed a higher retention (80%) compared to the

Table 2. Individual intraclass correlation coefficient for heart rate during the cycling protocol for Cart and Load Effort Rating (CALER) and Bug and Bag Effort (BABE) rating across Trial 1 (T1) to Trial 3 (T3)

	Cycling			Stepping		
	T1–T2	T2–T3	T1–T2–T3	T1–T2	T2–T3	T1–T2–T3
CALER 3	0.56*	0.76	0.69	0.75	0.75	0.73
CALER 5	0.77	0.76	0.73	0.86	0.80	0.82
CALER 8	0.74	0.80	0.72	0.81	0.84	0.84
Overall	0.74	0.81	0.74	0.84	0.85	0.83
BABE 3	0.71†	0.93	0.78	0.76	0.84	0.77
BABE 5	0.85	0.88	0.86	0.85	0.89	0.84
BABE 8	0.88	0.88	0.87	0.89	0.73	0.79
Overall	0.84	0.90	0.84	0.88	0.87	0.84

All $p < 0.001$ unless otherwise indicated. *$p < 0.05$; †$p < 0.01$.

Table 3. Limits of Agreement (LoA) analysis (bias ± 95% LoA) of heart rate for Cart and Load Effort Rating (CALER) and Bug and Bag Effort (BABE) rating during the cycling and stepping exercise modes across Trial 1 (T1) to Trial 3 (T3)

Effort level	Cycling				Stepping			
	CALER		BABE		CALER		BABE	
	T1–T2	T2–T3	T1–T2	T2–T3	T1–T2	T2–T3	T1–T2	T2–T3
3	5 ± 33	−4 ± 21	−6 ± 29	−1 ± 14	1 ± 19	2 ± 21	2 ± 19	−4 ± 17
5	7 ± 21	−4 ± 20	2 ± 20	−1 ± 18	1 ± 15	0 ± 21	1 ± 16	−4 ± 15
8	3 ± 18	−2 ± 15	1 ± 17	0 ± 16	1 ± 18	2 ± 18	2 ± 13	−4 ± 21
Overall LoA	5 ± 23	−3 ± 18	−1 ± 23	−1 ± 16	1 ± 17	2 ± 20	1 ± 16	−4 ± 17

CALER group (36%), with fewer children wishing to change from having previously used the BABE scale to using the CALER scale (20%). A significantly higher ($p < 0.05$) percentage of children wished to change from having previously used the CALER scale to using the BABE scale (64%).

Discussion

Results from this study show that children aged 7–11 years were able to regulate exercise efforts during intermittent cycle ergometry and stepping tasks (with randomized order of levels) by applying their understanding of the CALER or BABE scale. This study has also confirmed that the CALER and BABE child-specific effort rating scales may be used interchangeably. Results have shown that both scales provide similar consistencies within exercise modes, suggesting that either scale could be employed in children aged 7–11 years.

The increase in HR at RPE levels 3, 5 and 8 confirmed that the children understood the effort-rating scales used in this study. HRs were significantly higher for the cycling protocol than for the stepping protocol, irrespective of the scale used.

The ICCs showed a general trend of increasing reliability across trials. This concurs with previously reported effort production data during a cycling protocol (Eston et al. 2000) and demonstrates that the children had learned how to self-regulate exercise intensity. However, for the stepping protocol, the LoA analyses showed that practice did not enhance the children's production of exercise effort across trials. There is a lack of research on the reliability of children's effort production during intermittent stepping protocols. However, anomalous reliability results obtained from this study might suggest that children aged 7–11 years require more than three effort production trials during intermittent stepping protocols to achieve highly reliable results. Eston et al. (2000) highlighted the importance of three or more trials in the production of highly reliable exercise efforts in children aged 7–10 years during intermittent cycle ergometry tasks.

This study has demonstrated that the CALER and BABE scales, which are pictorially quite different, may be

applicable to a variety of exercise environments. In terms of reproducing a similar degree of physiologic strain, it appears that the two scales may be used interchangeably. The development of additional pictorial scales designed specifically to accommodate different modes of exercise in children, which appears to be the trend in the developments of the OMNI Scale (Robertson 2004; Utter et al. 2002), may therefore not be necessary. Results have demonstrated that the CALER and BABE scales are intermodal. Either scale may be used for cycling or stepping tasks with children aged 7–11 years.

With regard to preference of a particular scale, we had observed in previous validation work on the BABE and CALER scales (Eston et al. 2001) that children seemed to prefer the BABE scale, with many children disappointed not to be assigned to a BABE group in our studies. The BABE scale was developed to appeal to children, particularly as it contained characters that were similar to characters depicted in the then recent Walt Disney film "*A Bug's Life*", the central characters being familiar to most children. During informal discussions, it was indicated that reasons for their preferred choice were that the scale was "fun-looking" and "interesting" due to the use of colors, and the character association with a recent hit movie which all participants were familiar with. A more objective analysis of the children's preference revealed that most children would prefer to use the BABE scale rather than the CALER scale in future exercise sessions. Perhaps the BABE scale is also more pictorially representative of a range of ambulatory movements. However, both scales can be criticized on the basis of their linear nature. Indeed, all previous pictorial scales depict either a horizontal scale or a linear gradient. We are currently exploring the notion of using a pictorial curvilinear scale (Eston & Parfitt 2006). We believe that such a scale would have inherently obvious face validity as it is readily conceivable that a child will recognize from previous learning and experience that the steeper the hill, the harder it is to ascend. This assumes that the child is focused on the pictorial aspects of the scale, rather than the numbers or anchored phrases, and this focus may change with age. Future research could systematically explore the relative valence of the visual information (pictures, numbers, verbal cues, slope) available to the child.

References

Bland JM, Altman DG (1986). Statistical methods for assessing. agreement between two methods of clinical measurement. *Lancet* i:307–10.

Borg G (1998). *Borg's Perceived Exertion and Pain Scales*. Human Kinetics, Champaign, IL.

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

Eston RG, Lamb KL, Bain A, Williams M, Williams JG (1994). Validity of a perceived exertion scale for children: a pilot study. *Percept Mot Skills* 78:691–7.

Eston RG, Lamb KL (2000). Effort perception. In: Armstrong N, Van-Mechelen W (eds). *Paediatric Exercise Science and Medicine*. Oxford University Press, Oxford, pp 85–91.

Eston RG, Parfitt G, Campbell L, Lamb KL (2000). Reliability of effort perception for regulating exercise intensity in children using a Cart and Load Effort Rating (CALER) Scale. *Pediatr Exerc Sci* 12:388–97.

Eston RG, Parfitt G, Shepherd P (2001). Effort perception in children: implications for validity and reliability. In: Papaionnou A, Goudas M, Theodorakis Y, (eds). *Proceedings of 10th World Congress of Sport Psychology*, Skiathos, Greece, Vol. 5, pp 104–6.

Eston RG, Parfitt G (2006). Effort perception. In: Armstrong N (ed). *Paediatric Exercise Physiology*. Elsevier, Amsterdam, in press.

Lamb KL (1996). Exercise regulation during cycle ergometry using the CERT and RPE scales. *Pediatr Exerc Sci* 8:337–50.

Lamb KL, Trask S, Eston RG (1997). Effort perception in children: a focus on testing methodology. In: Armstrong N, Kirby BJ, Welsman JR (eds). *Children and Exercise XIX Promoting Health and Well Being*. E & FN Spon, London, pp 258–26.

Leung ML, Cheung PK, Leung RW (2002). An assessment of the validity and reliability of two perceived exertion rating scales among Hong Kong children. *Percept Mot Skills* 95:1047–62.

Robertson RJ, Goss FL, Boer NF (2000). Children's OMNI Scale of perceived exertion: mixed gender and race validation. *Med Sci Sports Exerc* 32:452–8.

Robertson RJ, Goss FL, Bell JA, Dixon CB, Gallagher KI, Lagally KM, Timmer JM, Abt KL, Gallagher JD, Thompkins T (2002). Self-regulated cycling using the children's OMNI Scale of Perceived Exertion. *Med Sci Sports Exerc* 34:1168–75.

Robertson RJ (2004). *Perceived Exertion for Practitioners: Rating Effort with the OMNI Picture System*. Human Kinetics, Champaign, IL.

Utter AC, Robertson RJ, Nieman DC, Kang J (2002). Children's OMNI Scale of perceived exertion: walking/running evaluation. *Med Sci Sports Exerc* 34:139–44.

Williams JG, Eston RG, Furlong B (1994). CERT: a perceived exertion scale for young children. *Percept Mot Skills* 79:1451–8.