EVIDENCE OF CONTENT, CONCURRENT CRITERION VALIDITY, AND RELIABILITY IN A *CHASSÈ* CHECKLIST

NÁDIA F.S. MARINHO[1,2]**, LAYLA M.C. ABURACHID**[3]**, THÁBATA V.B. GOMES**[4]**,
PATRICK C. RIBEIRO-SILVA**[1,5]**, MARCELO G. DUARTE**[1,6]**, ISA H.G. FLOR**[2]**,
CARINA A.S. ROMANA**[1]**, PABLO J. GRECO**[7]**, RODOLFO N. BENDA**[8]

[1] Department of Physical Education, School of Physical Education, Physiotherapy and Occupational Therapy, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[2] Department of Physical Education, Universidade Salgado de Oliveira, Belo Horizonte, Brazil

[3] Department of Physical Education, Universidade Federal do Mato Grosso, Cuiabá, Brazil

[4] Department of Sport, Escola Superior de Educação Física, Universidade Federal de Pelotas, Pelotas, Brazil

[5] Health Center, Fundação de Ensino e Pesquisa do Sul de Minas- Centro Universitário do Sul de Minas, Varginha, Brazil

[6] Instituto de Ciências Sociais, Educação e Zootecnia, Universidade Federal do Amazonas, Manaus, Brazil

[7] Department of Sports, School of Physical Education, Physiotherapy and Occupational Therapy, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[8] Department of Gymnastic and Health, Escola Superior de Educação Física, Universidade Federal de Pelotas, Pelotas, Brazil

ABSTRACT

Purpose. This study aims to verify the evidence of content, concurrent criterion validity, and reliability in a *chassè* checklist.
Methods. First, 3 judges measured 11 items of the checklist based on the Likert scale, varying from 1 (not at all) to 5 (very much clear, pertinent, and representative), for the content evidence validity. In the second stage, 91 girls (12 ± 2.09 years old) were evaluated while performing the skill; 51 of them were skilled and 40 were novice in polka type *chassè* (concurrent evidence). Concerning reliability, 3 evaluators analysed the videos of 5 volunteers while performing polka type *chassè* at 2 different times, with an interval of 1 week between evaluations.
Results. After the first stage, 1 item presented the content validity coefficient of 0.56 and it was reformulated as well as the skill was renamed to polka type *chassè*. The second assessment presented clear, pertinent, and representative items of the instrument. The reliability was considered appropriate by the inter-judge intraclass correlation coefficient (ICC) of 0.83 and the lowest intra-judge ICC of 0.80. Lastly, concurrent evidence was better when it came to skilled participants than novice ones related to polka type *chassè*.
Conclusions. In general, the results indicated content and concurrent validity and reliability evidence. The instrument can be useful for physical education teachers as well as gymnastics coaches.
Key words: instrumentation, psychometrics, validity and reliability evidence, sports skill

Introduction

Studies of motor control, learning, and development are integrated in the motor behaviour area. In order to answer some questions on motor behaviour, different kinds of motor skills in distinct contexts were investigated [1]. Newell and Rovegno [2] and Newell [3] pointed out the need for motor behaviour research to go beyond the simple skills, limited to the laboratory settings. One of the goals of the motor behaviour area is to understand motor learning and development as well as to provide recommendations for physical education teachers and sports coaches. Thus, it might be important to investigate motor learning and development using non-laboratory tasks, closer to real conditions of teaching or training, which is generally a challenge [4].

This approach requires the use of instruments to obtain data that allow access to changes in performance or identify distinct performances between subjects in different learning stages, which, in turn, requires a validation process and reliability analysis. The reliability of an instrument refers to its consistency or repeatability and can be demonstrated by analysing its objectivity [5]. The validity of an instrument refers to the degrees in which the results obtained in it are read on the basis of data and a theory, enabling its use to what the instrument was proposed [6]. Therefore, this study surveyed existing instruments described in literature that address the assessment of specialized skills to check the existence and characteristics of these instruments.

In the bibliographical survey in databases as LILACS, SciELO, Medline, and Scopus, the use of the terms 'instrument,' 'validity,' 'validation,' 'validity evidence,' 'motor,' 'technique,' 'sport,' and 'specialized skill' allowed to find an instrument that assesses the pattern of free throw in basketball [7] and one that assesses the pattern of passing in volleyball [8]. In both cases, the analysis of content validity was performed from the instrument as a whole, which might dissemble it regarding pertinence of practice, language clarity, and theoretical representativeness of each item for the proposed analysis of the pattern.

The validations constitute a process in which different kinds of evidence accumulate and enable the understanding of an instrument's scores in accordance with a finality to which it was designed. For example, there is evidence based on (a) content; (b) response processes; (c) the internal structure of tests; and (d) relations to other variables. Evidence based on relations to other variables is classified into convergent and discriminant validity evidence as well as criterion validity evidence (concurrent and predictive) [6]. The following evidence was investigated in this study: (a) validity evidence based on content; (b) evidence based on relations to other variables, such as concurrent criterion-related evidence; and (c) reliability evidence.

The first step is the theoretical definition of a construct to be assessed, that is, a finite universe of behaviours that will be assessed by the instrument. Then, there is a construction and search of content validity evidence that represents this construct. Content validity is the last step of the theoretical procedures to be reached. This one consists in checking if a test represents the construct in a clear, representative, and pertinent way [9].

This study investigated the construct of *chassè*, as there has not been found, in the current literature, an instrument that would assess the pattern of this skill. *Chassè* is a term that designates a gliding step in which the back foot moves the front one as if one were chasing the other [10]. It becomes noticeable that the definitions found refer to operational definitions, particularly, guided by experience and observation [9] of the skill pattern in *chassè*. Theoretically, the feature to be assessed in *chassè* is its pattern, which may present variations between ballet and gymnastics. This specialized pattern has dimensions that refer to the moves of legs, feet, arms, and hands.

The relation of the test scores and an external relevant criterion related to it might guide another source of validity (concurrent criterion-related) evidence. This kind of concurrent criterion-related validity evidence is based on the simultaneous use of test scores and the external criteria outcome that must be their good predictor [6]. The distinct performance of a skilled person in comparison with a novice one might be an external criterion of an instrument that intends to identify differences related to performance in a motor pattern.

In this context, the aim of this study was to verify the validity and reliability evidence of an instrument utilized to assess the *chassè*. This instrument may help researchers who want to follow a more ecological approach, using more field-based tasks in sports research [4].

Material and methods

This is a methodologic study [9], and its description and results are oriented by the kinds of evidence of validity obtained in the sequential order: content validity, concurrent criterion-related evidence, and reliability evidence.

Content validity evidence

Participants

Ten judges were invited to take part in this study, and only 4 of them returned their analyses. After tabulating data and identifying items that were not answered by the judges, only the 3 remaining results were included in the study. The criteria of the judges selection involved a minimum of 10-year experience in the area [11], taking into account the acquisition of this experience via deliberate practice [12]. Besides, the participant judges were referees or rhythmic gymnastics coaches, positioned by athletes in the first 8 places in the current or past national championships; moreover, they were working as coaches or as referees at a na-

Table 1. Judges' features

Specialist	Years of experience	Area of business	Titles/workstream
1	34 years	Coach	Currently has an athlete in the Brazilian Gymnastics Group
	16 years	Referee for the International Federation of Gymnastics	Referee in international championships, including the Olympics in Rio 2016
2	16 years	Coach	Athletes among the 8 best of the state in youth teams
3	27 years	Coach	3 athletes among the 11 best in national rankings, 3 athletes among the 8 best in Pan Americano, and 1 athlete at the 4th place in the International Gymnastics Federation 2017 world ranking

tional and/or international level. The average experience was 19.5 years (± 14.4) (Table 1).

Procedures

Content validity implies the representativeness of a finite universe of behaviours (constructs); for this study, the execution from a psychometric model of instrumental elaboration by Pasquali [9] was selected. The checklist items were initially elaborated from the description of the skill components and then presented to a national rhythmic gymnastics referee and professor with 25 years of experience in this modality. The initial version involved 14 items, divided into elements that assess the patterns of movements of arms (4 items), legs (6 items), hands (2 items), and feet (2 items).

After the initial assessment, 4 items were rewritten, 3 items were eliminated, and 11 items remained: 5 to assess leg movements, 1 to assess arms, 3 to assess feet, and 2 items to assess hands. This version was then sent to the judges.

For content evaluation of the *chassè* checklist, the judges assessed each item for language clarity (does this item describe the pattern of movement that makes comprehension easy?), pertinence of practice (does this item properly represent an element of a polka type *chassè* skill?), and theoretical representativeness (does this item assess the polka type *chassè* skill?) [13]. The evaluations by the judges were measured on a 5-point Likert scale for each criterion, varying from 1 (not at all) to 5 (very much clear, pertinent, and representative). The judges could also include their suggestions each time they assessed an item in 3 points or less or when they considered necessary the inclusion or exclusion of any information.

Analysis

For the calculation of content validity, the authors selected the content validity coefficient (CVC) by Hernández-Nieto [13]. The first step was to calculate the average scores given by the judges to each item for each criterion (language clarity, pertinence of practice, and theoretical representativeness) from the formula below:

$$Mx = (\Sigma xi)/J$$

In the formula, *Mx* represents the average scores given by the judges, Σxi is the sum of the scores given by the judges, and *J* stands for the number of judges. Later, the CVC of each item (*CVCi*) was calculated, considering *Vmax* the maximum score that could be given to each item:

$$CVCi = Mx/Vmax$$

The calculation of content validity proposed by Hernández-Nieto [13] considers the possibility of bias in the assessment performed by the judges; the following formula allows to calculate the mistake (Pe_i):

$$Pe_i = (1/J)^J$$

The CVC of each item taking the value of the constant out (*CVCc*) and the total CVC (CVC_t) can be calculated by the use of the following formulas:

$$CVCc = CVCi - Pe_i$$
$$CVC_t = Mcvc_i - Mpe_i$$

$Mcvc_i$ represents the average of CVCs from the items and Mpe_i the average of mistakes in each item in the test. Total CVC (CVC_t) refers to the calculated value for the instrument in general, which allows the assessment in which the items represent a construct [14].

Hernández-Nieto [13] and Pasquali [9] recommend that the items of an instrument must present a value of CVC ≥ 0.80. However, Hernández-Nieto [13] points out that the levels of CVC higher than 0.70 are considered acceptable.

Concurrent criterion-related validity and reliability evidence

Participants

Overall, 91 girls took part in this study; 51 were skilled (12.7 years old ± 2.02), presenting at least 2 years of practice in rhythmic gymnastics and 1 year after learning the *chassè* skill. They had practised *chassè* at least twice a week for 1 year, as well as they had competed at least at the municipal level. Forty girls were novice (12.2 years old ± 2.12) in this task: they had never practised *chassè* before. The novice girls were athletes of a basketball team.

Procedures

First, the free consent terms were handed out and explained to the parents of the rhythmic gymnastics athletes and novice girls. Then, the participants were asked to fill out the consent form.

The volunteers were then videorecorded when performing 2 trials of the polka type *chassè* skills. Later, 3 judges analysed the videos in 2 moments, with a 7-day interval between them. The polka type *chassè* assessment was conducted with a checklist previously analysed in terms of content validity evidence. The performance of polka type *chassè* (Appendix) was rated 1 point when it met the criterion of each item and 0 when the performance was incorrect. There were 11 items and the task was performed twice. The lowest possible result on this skill was 0 and the highest equalled 22 points.

Analysis

The analysis of the criterion-related validity evidence (concurrent) involved the Mann-Whitney U test comparing skilled and novice groups with reference to their polka type *chassè* performance. In order to analyse reliability, the recorded movements of 5 volunteers performing the polka type *chassè* were assessed by 3 trained evaluators [15]. The analyses were performed at 2 different times, with an interval of 1 week between them. These data served to analyse the consistency of the measurements performed by using the intraclass correlation coefficient (ICC) [14]. After obtaining a satisfactory ICC (above 0.70) [16] in these stages, the other videos were randomly distributed among the evaluators; each evaluator was responsible for analysing a portion of the volunteers.

Ethical approval

The research related to human use has complied with all the relevant national regulations and institutional policies, has followed the tenets of the Declaration of Helsinki, and has been approved by the Internal Review Board of the Federal University of Minas Gerais (ETIC 2.081.647).

Informed consent

Informed consent has been obtained from all individuals included in this study.

Results

Content validity evidence

After the first stage of assessment, it was observed that although the total CVC of the instrument was 0.86 (Table 2), item 4 presented unsatisfactory CVC for the enquiry of pertinence of practice (Table 3). The suggestions presented by the judges indicated that it would be interesting to include a picture of the skill in the instrument and name the kind of *chassè* assessed. Besides, the term 'fall' was inadequate for this modality. A new round of assessment was performed after introducing the changes proposed. As the CVC formulated by Hernández-Nieto [13] allows a new assessment by the judges regarding the image or language clarity, the authors of this study understood that after the inclusion of the image, the instrument should be submitted to a new assessment. Such action aided an increase in the coefficient of this and other criteria, as shown in the following results.

All criteria, in their total values, obtained a start value of 0.88 (CVC_t) in their first stage of assessment (Table 2) and the CVC_t of the instrument was 0.86. Item 4, which obtained unsatisfactory assessment regarding pertinence ($CVCi$ = 0.56) in the first assessment (Table 3), caused changes in the instrument and it was reassessed by the same judges, as described in the previous paragraph. In the second assessment, the changes in the instrument were considered clear, pertinent, and representative for the skill. At last, total CVC for each criterion in the second assessment reached a start value of 0.89 and the CVC_t of the instrument turned to 0.88.

Table 2. Content validity coefficient for the criteria of clarity, pertinence, and representativeness

Validity criteria	CVC_t first assessment	$CVCc$ reassessed item		CVC_t after reassessment
		1st time	2nd time	
Clarity	0.88	0.82	0.89	0.89
Pertinence	0.89	0.56	0.89	0.92
Representativeness	0.92	0.76	0.89	0.93
Instrument	0.86	–	–	0.88

CVC – content validity coefficient, CVC_t – total CVC, $CVCc$ – constant CVC

Table 3. Content validity coefficient for the criteria of clarity, pertinence, and representativeness per item in 1st and 2nd time assessments

| Item	Validity criteria	Clarity		Pertinence		Representativeness	
		1st	2nd	1st	2nd	1st	2nd
1	Step ahead with right leg that propels a leap	0.76	–	0.76	–	0.89	-
2	Left leg gets closer or touches right leg in the air	0.89	–	0.89	–	0.96	–
3	Extended legs in the air	0.96	–	0.96	–	0.96	–
4	Fall with left leg	0.83	–	0.56	–	0.76	–
	Landing with left leg	–	0.89	–	0.96	–	0.96
5	Repeats sequence with left leg in front	0.89	–	0.89	–	0.89	–
6	Arms pass through 1st position (when on the ground) to the 2nd (when in the air)	0.82	–	0.96	–	0.96	–
7	Relevé position when on the ground	0.82	–	0.96	–	0.96	–
8	Tiptoes whenever foot is in the air	0.96	–	0.96	–	0.96	–
9	Tiptoes when foot lands on the ground	0.96	–	0.96	–	0.96	–
10	Hands position as a continuity of arms	0.89	–	0.96	–	0.96	–
11	Thumbs a little lower than other fingers	0.89	–	0.89	–	0.82	–
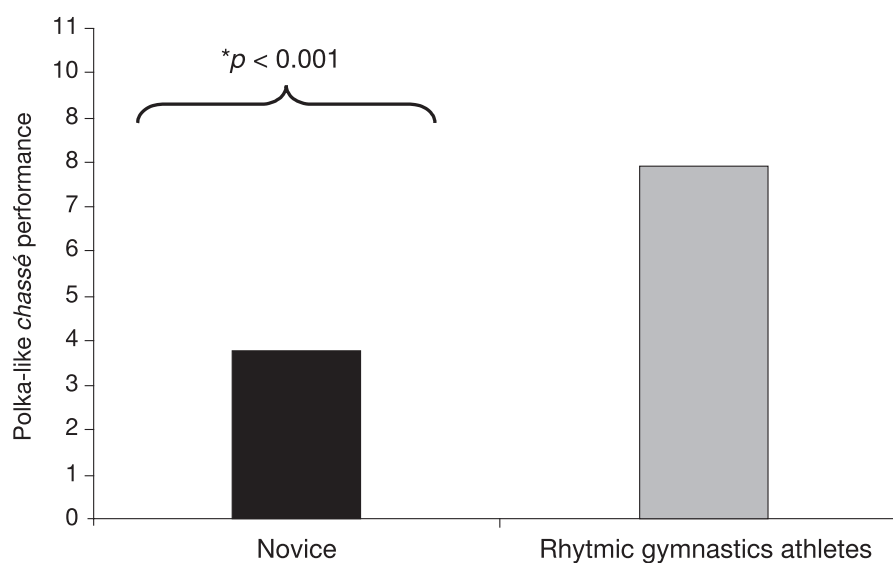
Figure 1. Comparison of average performance in 2 trials for polka-like *chassè* ($p < 0.001$)

Concurrent criterion-related validity
and reliability evidence

The analysis of reliability pointed at an inter-judge ICC of 0.83; the lowest intra-judge ICC of 0.80 indicated an appropriate level [16]. This finding allows to assume it as reliability evidence.

Regarding the analysis of criterion-related validity evidence, the inter-group comparison revealed that the group skilled in *chassè* presented a significantly higher performance ($p < 0.001$) than the novice group (Figure 1).

Discussion

This study analysed the content, concurrent criterion-related validity, and reliability evidence of an instrument utilized to assess polka type *chassè*. The results of the assessment by the judges indicated that the instrument presented language clarity, theoretical representativeness, and pertinence of practice, although 1 item showed unsatisfactory CVC.

The item 'fall with left leg' was not considered representative of the skill in the first assessment, that is, the item did not represent adequately an element of *chassè*. The suggestions elaborated by the judges indicated that there were more patterns than one for this component and that it would be interesting to insert an picture of the skill in the instrument and name it 'polka type *chassè*' in order to specify the kind of pattern utilized. Besides, the term 'fall' could denote a grave fault in the rhythmic gymnastics modality. So, we decided to rename this item to 'landing with left leg' along with a picture of the pattern, and the instrument was then called a 'checklist for the assessment of the polka type *chassè* skill' (Appendix).

These results point at the importance of analysing language clarity, theoretical representativeness, and pertinence of practice not utilizing the instrument in a general way, as performed by de Souza Nunes et al. [7] and Magarotto Junior and Deprá [8]. Valentini et al. [17] also applied the procedure of item analysis by evaluators when assessing validity of the content of the Test of Gross Motor Development-2 concerning the population of the Brazilian state of Rio Grande do Sul. A general analysis assumes that all the items are adequate or not in their totality [18] and might bias the results of a validation process and change results by the use of a mistaken instrument.

The results of intergroup comparisons also indicated that the polka type *chassè* checklist could be considered sensible to distinguish groups with different levels of skill, pointing at evidence of concurrent criterion-related validity [6]. In a study by de Souza Nunes et al. [7], 4 subjects presented different learning stages. The study aimed to identify kinds of errors typical of those different learning stages. The conclusion was that the novice participants made more errors than the skilled ones. These results point out that an instrument can differentiate subjects with different learning stages (external criteria) [6], being sensible to be used in motor learning studies.

This sort of evidence, established by an external criterion, allows, from the search of groups with specific features, the chance to check if the instrument is sensible to assess differences between subjects in the so-called construct proposed for this measurement [19]. Regarding the instrument proposed to assess the polka type *chassè*, it is important that it is sensible to detect subjects at different learning stages or that it can point at performance changes in the very skill referred to along the practice.

It should be highlighted that the instrument of the present study can contribute to future studies on the motor behaviour area [1], aiming at a wider ecological validity, since the skill is utilized in dance and gymnastics modalities [4].

Conclusions

The results obtained from the assessment by the judges indicate that the instrument, after going through a series of evaluation processes, presented content validity evidence. Furthermore, the instrument, based on the analysis of evidence and competing criteria, was sensitive to detect differences in performance between novice and skilled subjects.

A limitation of the study is that the instrument analysis was performed in a sample composed exclusively of girls. This condition is due to the fact that in gymnastics, as in ballet, some patterns differ in boys and in girls. In *chassè*, the position of fingers would be an example of this distinction.

Disclosure statement
No author has any financial interest or received any financial benefit from this research.

Conflict of interest

The authors state no conflict of interest.

References

1. Magill RA, Anderson DI. Motor learning and control: concepts and applications, 11th ed. New York: McGraw-Hill Education; 2017.
2. Newell KM, Rovegno I. Commentary – motor learning: theory and practice. Quest. 1990;42(2):184–192.
3. Newell KM. Change in motor learning: a coordination and control perspective. Motriz. 2003;9(1):1–6; doi: 10.5016/1134.
4. Merbah S, Meulemans T. Learning a motor skill: effects of blocked versus random practice. A review. Psychologica Belgica. 2011;51(1):15–48; doi: 10.5334/pb-51-1-15.
5. Thomas JR, Nelson JK, Silverman SJ. Research methods in physical activity. Champaign: Human Kinetics; 2005.
6. Yun J, Ulrich DA. Estimating measurement validity: a tutorial. Adapt Phys Activ Q. 2002;19(1):32–47; doi: 10.1123/apaq.19.1.32.
7. De Souza Nunes ME, Gehring PR, Basso L, Oliveira da Fonseca MC, Thomazi MG, Santos S. Development and validation of an instrument to assess movement quality of the basketball free throw shooting [in Portuguese]. Motriz. 2012;18(4):627–635; doi: 10.1590/S1980-65742012000400001.
8. Magarotto Junior LA, Deprá PP. Validating a checklist for the qualitative analysis of volleyball reception [in Portuguese]. Motriz. 2010;16(3):571–579; doi: 10.5016/1980-6574.2010v16n3p571.
9. Pasquali L. Psychometrics. Rev Esc Enferm USP. 2009;43(Special):992–999; doi: 10.1590/S0080-62342009000500002.
10. Faro AJ, Sampaio LP. Ballet and dance dictionary [in Portuguese]. Rio de Janeiro: Jorge Zahar; 1989.
11. Balbinotti MAA, Benetti C, Soares Terra PR. Translation and validation of the Graham-Harvey survey for the Brazilian context. Int J Manag Finance. 2007;3(1):26–48; doi: 10.1108/17439130710721644.
12. Ericsson KA, Nandagopal K, Roring RW. Toward a science of exceptional achievement: attaining superior performance through deliberate practice. Ann N Y Acad Sci. 2009;1172:199–217; doi: 10.1196/annals.1393.001.
13. Hernández-Nieto R. Contributions to statistical analysis. Mérida: Universidad de Los Andes; 2002.
14. Commenges D, Jacqmin H. The intraclass correlation coefficient: distribution-free definition and test. Biometrics. 1994;50(2):517–526; doi: 10.2307/2533395.
15. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–240; doi: 10.1519/15184.1.
16. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–428; doi: 10.1037//0033-2909.86.2.420.
17. Valentini NC, Barbosa MLL, Cini GV, Pick RK, Spessato BC, Balbinotti MAA. Test of Gross Motor Development: expert validity, confirmatory validity and internal consistence [in Portuguese]. Rev Bras Cineantropom Desempenho Hum. 2008;10(4):399–404; doi: 10.5007/1980-0037.2008v10n4p399.
18. Pasquali L, Primi R. Basic theory of item response theory – IRT [in Portuguese]. Aval Psicol. 2003;2(2):99–110.
19. Lima EP, Vasconcelos AG, Barreto SM, Assunção AÁ. Checklist of occupational traumatic events for emergency professionals: adaptation and validation [in Portuguese]. Aval Psicol. 2016;15(3):391–401; doi: 10.15689/ap.2016.1503.12.

APPENDIX

Checklist for the assessment of the polka type *chassè* skill

Score 1 was granted when the pattern performed was correct and 0 when the pattern performed was not correct for each item.

Item	Legs pattern	1st	2nd
1	Step ahead with right leg that propels a leap		
2	Left leg gets closer or touches right leg in the air		
3	Extended legs in the air		
4	Fall with left leg		
5	Repeats sequence with left leg in front		
	Arms pattern		
6	Arms pass through 1st position (when on the ground) to the 2nd (when in the air)		
	Feet pattern		
7	Relevé position when on the ground		
8	Tiptoes whenever foot is in the air		
9	Tiptoes when foot lands on the ground		
	Hands pattern		
10	Hands position as a continuity of arms		
11	Thumbs a little lower than other fingers		
	SUM OF ITEMS PER TRIAL		