

儿童动作发展测试(TGMD-2)信度和效度的研究

李 静, 马红霞

(山东师范大学 体育学院, 山东 济南 250014)

摘 要: 以 511 名 3~10 岁儿童为测试对象, 探讨了大肌肉动作发展测试 (TGMD-2) 的信度和效度, 结果表明测试者信度系数为 0.436~0.876 ;内部一致性信度系数为 0.607~0.925 ;重测信度系数为 0.608~0.876 ;内容效度、结构效度、效标效度也都令人满意。TGMD-2 可作为 3~10 岁儿童大肌肉动作发展状况的测评工具。

关 键 词: 大肌肉动作发展; 信度; 效度; 儿童

中图分类号: G80-32 文献标识码: A 文章编号: 1006-7116(2007)03-0037-04

Study of the credibility and validity of the test of gross motor development of children

LI Jing, MA Hong-xia

(College of Physical Education and Sports, Shandong Normal University, Jinan 250014, China)

Abstract: Basing their test object on 511 children at the age of 3-10, the authors probed into the credibility and validity of the Test of Gross Motor Development (TGMD-2), and revealed the following findings: The credibility coefficient of the testees is between 0.436 and 0.876; the credibility coefficient of internal consistency is between 0.607 and 0.925; the credibility coefficient of retest is between 0.608 and 0.876; the content validity, structure validity and criterion validity are all satisfying. TGMD-2 can be used as a tool for evaluating the condition of gross motor development of children at the age of 3-10.

Key words: gross motor development; credibility; validity; children

在个体发育的早期, 动作发展是评价、诊断、监测个体身心发展状况的重要指标。儿童发展的早期成就主要体现在动作的发展上, 从 20 世纪初开始, 对早期发展的评估就多以动作为主要指标。动作的良好发展可以为个体认知、情绪和社会性等多方面的发展提供有利条件, 反之, 动作发展出现障碍, 个体的发展也会受到阻碍^[1]。大肌肉动作发展是动作发展的重要内容。大肌肉动作技能就是“由躯干、四肢等大肌肉群发力而产生的动作技能”。大肌肉动作发展包括在空间内移动身体的技能(位移技能)和控制物体的技能^[2]。这些动作技能是儿童在学前期和小学早期重点发展的动作, 也是为以后运动技能发展提供基础作用的大肌肉动作。在学前期和小学期间, 孩子的运动能力开始发展并渐趋成熟起来。身体的发育状况和孩子的活动经历在影响运动模式方面起到根本作用, 如果大肌肉动作发展方面的缺陷得不到识别和补救, 孩子就有可能终身都存在运动能力方面的问题。大肌肉动作发展测试(test of gross motor development, TGMD)

由美国密歇根州立大学 Dale A. Ulrich 博士编制的, 专门用于评估 3~10 岁儿童大肌肉动作发展状况的测评工具。该测试始编于 1985 年, 经过几年的实践后并结合有关专家们的建议, Ulrich 博士于 2000 年又进行了修订(称 TGMD-2)(TGMD-2 内容详见中国体育科技 2005 年第 2 期 41 卷)。修订后的测试在美国体育教学和研究中广泛应用, 并在多种文化环境下都证实有良好的信度和效度。TGMD-2 是一项标准严格、具有高度有效性和可靠性的测验项目。该测验由两部分构成, 分别测验了儿童身体移动的能力和控制物体的能力。每个测验由不同的测试动作组成, 每个测试动作测量的是儿童大肌肉动作发展的不同方面。它的测验样本是由美国 10 个州的 1 028 名儿童组成, 并根据性别、年龄、种族、居住地区进行了分类, 经研究证明具有很高的信度和效度^[3]。

TGMD-2 的测试内容包括位移运动部分和物体控制部分。位移运动部分包括跑步(Run)、立定跳远(Horizontal Jump)、单脚跳(Hop)、跨跳(Leap)、前滑步(Gallop)、侧滑步(Slide)

6 个动作；物体控制部分包括原地拍球(Stationary Dribble)、接球(Catch)、踢球(Kick)、击固定球(Striking a Stationary Ball)、上手投球(Overhand Throw)、地滚球(Underhand Roll) 6 个动作，总共 12 个测试动作^[4]。在我国台湾地区的儿童中应用，也得到了较好的信度和效度。但在大陆儿童中应用还没有相应的报告。为此，本文将大肌肉动作发展测试(TGMD-2)在大陆儿童中试测，研究其信度和效度，以探讨将其推广使用的可行性。

1 研究对象与方法

1.1 研究对象

由济南胜利大街小学、济南银座双语幼儿园、德州市光明街小学、德州市直属机关幼儿园 3~10 岁儿童组成，共 511 名。其中 3 岁 61 名(男 32 名，女 29 名)；4 岁 65 名(男 31 名，女 34 名)；5 岁 67 名(男 35 名，女 32 名)；6 岁有 57 名(男 30 名，女 27 名)；7 岁 69 名(男 35 名，女 34 名)；8 岁 64 名(男 34 名，女 30 名)；9 岁 68 名(男 35 名，女 33 名)；10 岁 60 名(男 28 名，女 32 名)。男生共 260 名，女生共 251 名。

1.2 研究方法

(1)测试工具：大肌肉动作发展测试(TGMD-2)由两部分构成，分别测试儿童身体移动的能力和物体控制的能力。每个部分的测试由不同的测试动作组成，每个测试动作测量了儿童大肌肉动作发展的不同方面。位移测试包括跑步、立定跳远、单脚跳、跨跳、前滑步、侧滑步 6 个动作；物体控制测试包括原地拍球、接球、踢球、击固定球、上手投球、地滚球 6 个动作，共 12 个动作^[4]。每个测试动作有 3~5 项不同的、具体的动作标准。这些动作标准可以界定儿童在完成某个动作时躯干和四肢配合是否协调、基础大肌肉动作技能的形成是否精确，是否达到自然、顺畅的动作技能模式。测试人员根据儿童完成动作的情况也就是符合动作标准的情况记录得分，符合一项动作标准得 1 分，不符合动作标准不得分，记为 0 分。

(2)测试过程：正式测试之前，随机抽取 30 人进行预测试，统一测试人员的操作方法和评分标准，解决测试过程中可能出现的问题。测试开始后，有两名测试人员进行记录评分，每个儿童进行两次测验动作的测试，以减少随机误差的出现。将两次动作的得分之和作为最后成绩。在测试过程中，要鼓励学生做最大效能的动作表现，并且尽量让每一个学生都有先测的机会。两周后，从已测对象中抽取 80 名(每个年龄段 10 名)进行重测，重测率为 15.65%。所有数据都在 SPSS for Windows 12.0 软件中进行处理和分析。

2 结果与分析

2.1 信度

测试的信度(reliability)又称可靠性，是指在相同测试条件下，对同一批受试者使用相同的测试手段，重复测试结果的一致程度。一个好的测试必须是稳定可靠的，只有具有充分可信度的测试才能测试到“真实”。常见用来评价信度的方法有评分者信度、内部一致性信度和重测信度^[5]。

(1)评分者信度：评分者信度(raters reliability)是指两个或数个评价者(裁判、教师、测试者等)对同一受试者进行同一测试时，所得测量结果的一致程度^[6]。在本次测试中，两名测试人员经过统一培训后，对 511 名被测对象的动作表现进行了独立的记录评分。将两名测试人员的评分运用皮尔逊(Person)相关系数进行比较，结果见表 1。由表 1 中的数据可以看出，不同测试者对 12 个测试动作的评分情况在 $\alpha=0.01$ 的水平上，有显著性的相关关系，也就是两个评分者的评分情况是基本一致的，说明 TGMD-2 的评分者信度值很高，可为测试评分人员提供有力的可信证据。

表 1 不同测试人员评分 ($\bar{x} \pm s$) 情况的相关分析

测试项目	第 1 个评分者	第 2 个评分者	相关系数
跑	3.34 ± 0.73	3.32 ± 0.74	0.677 ¹⁾
立定跳	2.85 ± 0.90	2.82 ± 0.92	0.634 ¹⁾
单脚跳	3.15 ± 1.28	3.11 ± 1.29	0.861 ¹⁾
跨跳	2.49 ± 0.73	2.52 ± 0.72	0.655 ¹⁾
前滑步	2.93 ± 1.02	2.91 ± 1.05	0.719 ¹⁾
侧滑步	3.22 ± 1.09	3.16 ± 1.10	0.775 ¹⁾
拍球	2.43 ± 1.25	2.54 ± 1.26	0.806 ¹⁾
踢球	3.42 ± 0.82	3.42 ± 0.79	0.436 ¹⁾
接球	2.05 ± 0.81	2.06 ± 0.86	0.663 ¹⁾
击固定球	2.41 ± 1.06	2.47 ± 1.09	0.536 ¹⁾
上手投球	2.05 ± 1.26	2.15 ± 1.25	0.724 ¹⁾
地滚球	2.42 ± 1.00	2.45 ± 1.00	0.621 ¹⁾

1)在 $\alpha=0.01$ 水平有显著性意义的相关系数

(2)内部一致性信度：内部一致性信度(internal-consistency reliability)是指在一天内受试者重复参加相同的测试所获得成绩(结果)的一致程度。为了检验 TGMD-2 的内部一致性信度，在本次测试中，每个测试动作都完成了两次，采用 Cronbach's α 系数来表示量表的内部一致性。

表 2 列出了 TGMD-2 中 12 个测试动作的内部一致性系数，单脚跳的系数为 0.925，踢球的系数为 0.709，击固定球的系数为 0.748，其他动作的信度系数在 0.766~0.892。根据信度系数的参考标准，当 $0.5 < \alpha < 0.7$ ，很可信(最常见)；当 $0.7 < \alpha < 0.9$ ，很可信(次常见)；当 $\alpha > 0.9$ ，十分可信。可见，这些信度系数值有很高的接受性，可以放心使用。

表2 测试动作的内部一致性信度系数

测试动作	<i>a</i> 系数	测试动作	<i>a</i> 系数
跑	0.807	拍球	0.892
立定跳	0.776	踢球	0.709
单脚跳	0.925	接球	0.796
跨跳	0.792	击固定球	0.748
前滑步	0.837	上手投球	0.840
侧滑步	0.873	地滚球	0.766

(3)重测信度:重测信度(test-retest reliability)又称为稳定性系数,是指数天后,对受试者重复进行同样的测试时,测试结果的一致程度。重测信度高,说明测试的结果一致、可靠,测试成绩稳定;反之,说明该测试不稳定,需要进行修订^[7]。

测试结束两周后,又从测试对象中随机抽取了80名孩子(每个年龄段10名)进行重测,重测率为15.65%。将80名孩子的两次测试结果用皮尔逊(Pearson)积差相关系数进行相关分析,相关系数见表3。表中的数据表明每个测试动作前后两次的测试成绩在 $\alpha=0.01$ 水平上有显著性意义,这些数值对统计测试分数树立了稳定的信心。

表3 测试动作的重测信度系数($\bar{x} \pm s$)

测试项目	第一次测试	第二次测试	相关系数
跑	6.98±1.16	6.96±1.32	0.938 ¹⁾
立定跳	5.31±1.71	5.50±1.65	0.946 ¹⁾
单脚跳	6.25±2.36	6.65±2.38	0.891 ¹⁾
跨跳	4.98±1.40	5.10±1.37	0.918 ¹⁾
前滑步	5.16±1.92	5.35±1.93	0.948 ¹⁾
侧滑步	6.11±2.16	6.29±2.32	0.940 ¹⁾
拍球	4.88±2.31	5.45±2.44	0.910 ¹⁾
踢球	6.28±1.64	6.54±1.35	0.871 ¹⁾
接球	3.86±1.54	4.06±1.58	0.936 ¹⁾
击固定球	4.33±1.95	4.56±2.09	0.915 ¹⁾
上手投球	3.53±2.27	3.56±2.46	0.948 ¹⁾
地滚球	4.08±1.29	4.04±1.52	0.908 ¹⁾

1)在 $\alpha=0.01$ 水平有显著性意义的相关系数

2.2 效度分析

测试的效度(validity)就是测试的有效性,指一个测试或测量工具能真实地测量出所要测量事物的程度。一次测试是否有效,主要看其是否准确地测量了它所要测量的东西。效度越高,表示测量结果越能显示出所要测量对象的真正特征。常用来评价效度的标准有内容效度、结构效度和效标效度^[7]。

(1)内容效度。测试的内容效度(content validity)指测量的内容或项目对总体内容的代表性程度,也就是测试能否充分代表所要测量的内容范围。内容效度的高低,取决于测试题目的代表性,要看选出的题目能否包含所测内容范围的主要方面,并使各方面题目比例适当^[8]。

1)测试项目的辨别力。项目的辨别力(discrimination)有时也称区分度,就是测试项目对于不同水平的被试者加以区分

的能力。具有良好辨别力的测试,实际水平高的被试者应得高分,水平低的被试者应得低分。它是评价测试是否有效的“指示器”。通常用辨别力指数来评价。根据上下各27%的标准将被试者分为高分组和低分组,得出各测试动作的辨别力指数见表4。在一般测试中, $D>0.30$ 项目就可以接受了^[7]。可见TGMD-2各测试动作的辨别力指数是可以接受的。

表4 各测试动作的辨别力指数

测试项目	高分组	低分组	<i>D</i>
跑	8.00	5.03	0.370
立定跳	7.68	3.62	0.510
单脚跳	9.03	3.16	0.587
跨跳	6.00	3.20	0.470
前滑步	8.00	3.24	0.595
侧滑步	8.00	3.59	0.551
拍球	7.73	1.83	0.738
踢球	8.00	4.99	0.376
接球	5.86	2.10	0.627
击固定球	7.10	2.60	0.450
上手投球	6.96	1.17	0.724
地滚球	7.13	2.83	0.538

2)测试项目的难度。项目的难度(difficulty)就是测验的难易程度,可以反映出测试动作对学生的适宜程度。一个测试的难度是由组成测验的各项目难度决定的。通常采用某个测试动作的平均分与该动作满分之比来计算难度。 P 值越大,说明难度越小, P 值越小,说明难度越大。Anastasi和Urbina^[7]认为平均难度大约占50%并占有相当大分散性,分配在15%~85%的项目一般被认为是可以接受的。

从表5测试动作的难度值看出,击固定球和上手投球的难度最大,分别为49%和53%,踢球、跑、跨跳的难度相对较低,分别为86%、83%和83%。总的来说,12个测试动作的难度值都在可以接受的范围内。

表5 测试动作的难度值

位移分测验	平均得分	难度值/%	物体控制分测验	平均得分	难度值/%
跑	6.67	83	拍球	4.98	62
立定跳	5.68	71	踢球	6.85	86
单脚跳	6.26	63	接球	4.11	69
跨跳	5.00	83	击固定球	4.88	49
前滑步	5.84	73	上手投球	4.21	53
侧滑步	6.38	80	地滚球	4.87	61

(2)结构效度。结构效度(construct validity)又称构想效度,是指测验能够测量到理论上的结构或特质的程度。确定结构效度的方法可以通过分析测试内容与测试总分的相关关系来进行^[9]。表6列出了各个测试动作得分与测试总分的相关关系,其相关系数为0.421~0.719,差异均有显著性。说明各个测试动作与测试总分之间有密切的相互关系,测量表的结构合理。

表 6 各测试动作与总分的相关关系

测试项目	相关系数	测试项目	相关系数
跑	0.477 ¹⁾	拍球	0.719 ¹⁾
立定跳	0.573 ¹⁾	踢球	0.512 ¹⁾
单脚跳	0.712 ¹⁾	接球	0.672 ¹⁾
跨跳	0.454 ¹⁾	击固定球	0.589 ¹⁾
前滑步	0.421 ¹⁾	上手投球	0.619 ¹⁾
测滑步	0.641 ¹⁾	地滚球	0.573 ¹⁾

1)表示在 $\alpha=0.01$ 水平上存在显著性差异

(3)效标效度。效标效度是指测试分数能否有效的预测或估计某种行为表现。也就是一种测量的结果,与某种已证明有效的效标测量结果的相关程度^[9]。当有效性系数接近 1 时,说明该测量手段有效程度高;如果有效性系数接近 0,则说明该测量手段有效程度低。

由于目前我国还没有一套评价儿童大肌肉动作发展情况的量表,本文对于测定效标效度的“效标”就采用学生体育课的运动成绩及体育教师日常的评价。抽取德州市光明小学三年级一班 40 名学生 12 项测试动作的总分与他们学期末的体育运动成绩进行比较。两个成绩经过标准分的转换后进行相关分析,效标间的相关系数为 0.825, $P<0.01$, 可以认为 TGMD-2 的效标效度很高,可以接受。同时在测试过程中,通过与体育教师的交流了解到,测试结果与学生平日的体育成绩,以及有经验教师的经验评定是非常接近的,这也说明 TGMD-2 的测试结果经受了体育教育工作者的实践检验。

3 结论

(1)大肌肉动作发展是儿童最早发展起来的运动技能,它对儿童的心理发展和社会交往能力发展都有重要的作用。Ulrich 博士编订的大肌肉动作发展测验是专门用来评估 3 到 10 岁儿童大肌肉动作发展能力的量表,在多种文化环境下都证实有良好的信度和效度,并广泛应用于西方国家。因此,在我国研究大肌肉动作发展测验的可行性是非常必要的。

(2)对我国 511 名儿童的测验发现:位移分测验测试动作的得分情况比物体控制分测验测试动作要好,其中跑、侧滑步、踢球动作的掌握情况较好,击固定球和上手投球动作掌握情况相对较差。

(3)儿童大肌肉动作的发展能力随年龄的增长不断提高,不同的测试动作在不同年龄段表现出各自的特点。在位移分测验中不存在性别差异,在物体控制分测验中的踢球、击固定球等动作上存在非常显著的性别差异,男生比女生的动作得分要高,但从总体情况来看不存在性别差异。

(4)TGMD-2 有良好的信度:不同测试人员之间的信度系数为 0.860~0.922;不同测试动作之间的内部一致性信度系

数为 0.709~0.925,不同性别的内部一致性信度系数为 0.920~0.962,不同年龄的内部一致性信度系数为 0.755~0.930;重测信度系数为 0.871~0.948。不论是测试者信度、内部一致性信度还是重测信度都符合测量学的要求。

(5)TGMD-2 有良好的效度:专家评价以及对测试动作的区分度和难度的分析说明内容效度良好;测试动作之间以及测试动作与总分之间的相关关系和因素分析证明结构效度良好;以学生的运动成绩和教师的经验评价为效标也证明效标效度良好。

(6)TGMD-2 的测试内容丰富,量表长度适中,操作简单,易于掌握,是可靠有效方便可行的测试工具,可以在我国进行推广应用。

本文只是对 TGMD-2 的信度和效度进行了检验, TGMD-2 在中国广泛应用尚需要一段时间,因为还有大量的研究需要继续完成,比如扩大测试的样本,最好能加入农村和不同民族的样本;修订测验中的个别动作和动作标准,以适应中国儿童的实际情况;建立中国地区的常模,用来评价我国儿童的大肌肉动作发展状况;利用此量表致力于对智障儿童的研究。总之, TGMD-2 简单易行,信度和效度较好,是评价 3~10 岁儿童大肌肉动作发展状况的有效工具,值得推广应用。

参考文献:

- [1] Gregory V, Payne L, Isaacs D. Human motor development: A lifespan Approach 6th ed [M]. Boston: McGraw_Hill, 2005: 441.
- [2] Clark J E, Whitall J. What is motor development: The lessons of history [J]. Quest, 1990, 41: 183-202.
- [3] 李静,梁国力.大肌肉群发展测试(TGMD-2)研究[J].中国体育科技, 2005, 41(2): 105-107.
- [4] Dale A. Ulrich. Test of Gross Motor Development (Second Edition) Examiner's Manual [M]. Austin, TX: pro-ed Publishers, 2000: 3.
- [5] 施俊琦,王垒.一般性自我效能量表的信效度检验[J].中国心理卫生杂志, 2005, 19(3): 191-193.
- [6] 刘贤臣,郭传琴. Achenbach 青少年行为自评量表的信度和效度研究[J].中国心理卫生杂志, 1997, 11(4): 200-203.
- [7] 李志辉,罗平.SPSS for Windows 统计分析教程(第2版)[M].北京:电子工业出版社, 2005: 82.
- [8] 陈建华,蔡巧红.SPSS 软件在体育测量评价中的应用[J].体育科学研究, 2005, 19(1): 82-84.

[编辑: 郑植友]