**探究反应时作为因变量的优越性**

**摘 要** 反应时指从刺激呈现到个体做出明显反应所需要的时间，可以用来考察个体的心理加工过程。反应时具有独特的优越性，但也有一些研究指出它的不足。本实验采用线段区分范式，使用单一自变量、三个因变量的组内设计，比较被试的反应时、自信度和判断比较刺激比标准刺激长的比例（判断比例）的信度、效度、敏感性。结果表明，反应时的信度低于自信度和判断比例，三者均有较高的效度，判断比例出现高限效应和低限效应，反应时和自信度并未出现。结论为反应时优于判断比例，但无法说明反应时优于自信度。

**关键词**  反应时自信度 判断比例 线段区分

1 前言

反应时(Reaction Time, RT)指从刺激呈现到个体做出明显反应所需要的时间（朱滢，2014）。在许多研究当中，反应时被用来量化所研究的心理过程，例如思维旋转任务、注意偏向任务、内隐联想任务、以及Stroop任务等研究，都是通过比较反应时的差异来考察个体的心理加工过程（郭磊 等，2017）。

选取反应时作为因变量具有以下几个在功能上的优势。第一，在时间有限制的测验中，能够区分个体的速度和能力，如果只考虑作答的正确率，对于错答和未答的题目，就很难区分个体究竟是能力不足还是速度较慢，在测验中加入反应时可以很好地解决这一问题(Chuah & Zhang, 2007)。第二，有利于区分某些异常的反应模式（郭磊 等，2017）。第三，有助于比较个体在测验中采用的答题策略，例如在思维旋转测验中，一些个体可能采用分析规则策略，另一些可能采用思维旋转策略，尽管两种策略都能导致相同的反应结果，但它们的反应时的分布不同(Mislervy et al., 1991)。

反应时作为因变量具有较高的信度、效度和敏感性。Eide等（2002）用Stroop效应考察了反应时的信度，结果表明，不同条件下的前后测的反应时的一致性很高。反应时与工作记忆、情节记、结晶能力、感知速度等认知功能呈现显著的负相关，体现了反应时的效度(Hultsc et al., 2002)。Anstey等（1997）的一项知觉研究发现，反应时具有较高的敏感性。

自信度指个体相信自己的选择是正确或者最优的程度，且在决策中个体的自信度与选择的正确率呈现高度的正相关（余柳涛 等，2016）。Hemmon(1911)设计了线条区分范式来探究反应时与个体自信度的关系，要求被试完成选出两条线段中更长/短的一条，并在选择之后对准确性的自信度进行打分，结果表明，随着自信度的降低，反应时增加。延用线条区分范式，赫葆源等（1983）研究发现当个体在报告自信度为100%时，其反应时随着刺激差异的增加仍然在减少。这个研究结果说明反应时的敏感性比自信度更好，当自信度产生高限效应时，自变量的变化依然能够影响反应时。判断比较刺激比标准刺激更长的比例（判断比例）也可以作为因变量，它可以反映被试判断的正确率（当比较刺激更长时为正确率，当比较刺激更短时为错误率），有研究指出，正确率在刺激差异最小处变化最快(Treutwein & Strasburger, 1999)。

基于上述内容，本实验采用经典的线条区分范式，使用单因素组内设计，以比较刺激长度为自变量，以反应时、自信度、判断比例为因变量，探究反应时作为因变量的优越性。预期结果为反应时的信度、效度和敏感性均高于自信度和判断比例。

2 方法

**2.1 被试**

被试为75名北京大学参加实验心理学实验课程的学生，年龄在18到24岁之间（*M*=19.83, *SD*=1.19），男性32位，女性43位，视力或矫正视力正常。实验无报酬。

**2.2 仪器和材料**

本实验使用显示器为20-in ViewSonic，分辨率为1980×1020，刷新频率为59Hz，操作系统为Windows7的电脑。实验刺激为分隔开的两段线段，一段长为50mm，另一端则从40-60mm范围里变化，间隔2mm，共11个比较刺激。刺激材料如图1所示。

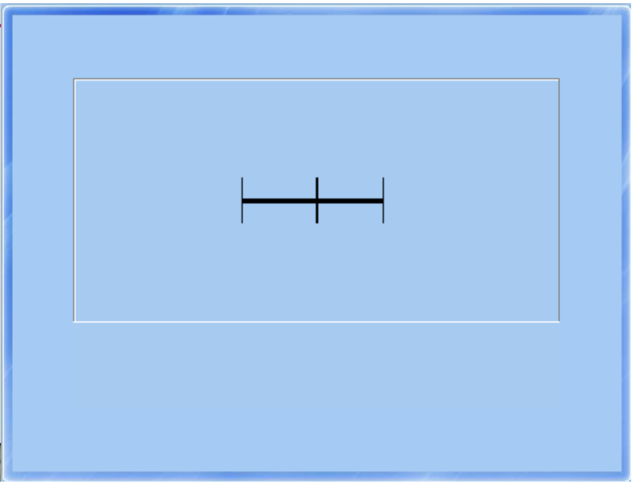


图1 刺激材料示意

**2.3 实验设计**

本实验采用单因素组内设计。自变量为比较线段的长度，范围从40mm到60mm，间隔为2mm，共11个水平。因变量有三个，分别为被试判断的反应时，被试判断的自信程度，被试判断比较刺激比标准刺激长的比例（判断比例）。反应时指从刺激呈现到被试反应的开始之间的时距，由电脑程序自动记录。被试的自信程度即被试对自己的判断有多大把握，从0（完全不肯定）到±100（完全肯定），由被试在每次判断后进行选择。被试判断比较刺激比标准刺激长的比例的操作性定义为对每一个比较刺激，被试判断该比较刺激比标准刺激长的次数除以10（每个比较刺激出现10次）。

本实验对潜在的额外变量进行了控制。首先，为了避免被试左右视野差异和习惯误差对实验结果的影响，实验中每种长度的比较刺激一半呈现在左侧，一半呈现在右侧，随机出现。其次，为了避免顺序效应，不同长度的比较刺激随机呈现。

**2.4 实验程序**

被试端坐在电脑前，开始实验。在显示器屏幕中央呈现一条粗的黑色水平线，两端各有一条短垂直线，另一条短垂直线将水平线分为两段。其中一段为50mm，作为标准刺激；另一段是比较刺激，长度范围从40mm到60mm，变化间隔为2mm，共有11个比较刺激。11个比较刺激随机呈现，每个比较刺激共呈现10次。其中5次呈现在左边，5次呈现在右边。在一个试次中，呈现水平线，被试判断是左边线段更长还是右边线段更长，电脑记录被试的反应时（单位为mm），每次判断后，屏幕下方会出现一个滑动条，被试可以用鼠标拖动滑动指针在滑动条上移动。不同的位置表示被试对自己正确反应的不同程度的信心，中间位置表示完全没有信心（0），最左端表示完全肯定左边线段长些（-100），最右端表示完全肯定右边线段长些（100），中间各点表示肯定的各种程度。一个试次的流程如图1所示。试次之间的间隔为2秒。

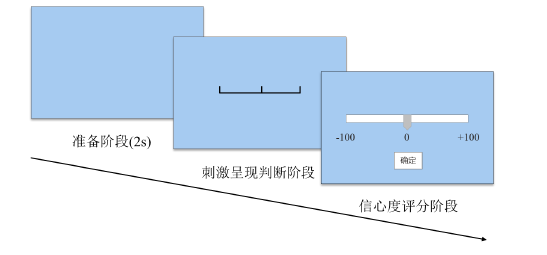


图2 一个试次流程示意

3 结果

使用SPSS 22进行数据分析，剔除三个标准差之外的数据，对余下数据做描述性统计，如表1所示。

表1 不同比较刺激长度下的反应时、自信度、判断比例(M±SD)

|  |  |  |  |
| --- | --- | --- | --- |
| 比较刺激长度/mm | 反应时/ms | 自信度 | 判断比例 |
| 40 | 781.42±287.48 | -96.37±7.58 | 0.00%±0.00% |
| 42 | 836.08±320.13 | -90.23±12.59 | 0.00%±0.00% |
| 44 | 930.24±397.56 | -82.30±15.73 | 0.00%±0.00% |
| 46 | 1163.84±493.73 | -57.68±26.48 | 4.59%±13.26% |
| 48 | 1694.86±873.40 | -26.14±19.65 | 19.08%±14.98% |
| 50 | 1796.85±950.94 | 0.37±10.45 | 51.37%±12.32% |
| 52 | 1513.35±711.94 | 33.18±19.94 | 87.11%±15.22% |
| 54 | 1115.06±453.94 | 58.10±24.89 | 99.68%±1.78% |
| 56 | 931.17±381.42 | 78.07±23.95 | 100.00%±0.00% |
| 58 | 836.95±341.98 | 90.40±14.52 | 100.00%±0.00% |
| 60 | 783.07±333.28 | 95.65±7.47 | 100.00%±0.00% |

以比较刺激长度为横坐标，以反应时，自信度、判断比例为纵坐标绘制折线图，如图3。从图中可以看出，被试的反应时以比较刺激长度为50mm为界，呈现倒U型，在50mm处达到最大值；自信度随着比较刺激增长平缓上升；判断比例呈现S型，在44mm、42mm、40mm均为0，在54mm、56mm、58mm、60mm均为1。

对不同比较刺激长度下的反应时、自信度和判断比例做正态性检验，结果显示，除42mm的自信度（*p*<.001），48mm的判断比例（*p*=.002），58mm的自信度（*p*<.001）外均满足正态性，由于本实验样本量较大（n=75），可以继续进行检验。

首先检验因变量的信度，由于本实验只进行了一次，所以使用分半信度来衡量因变量的信度。将被试按照编号的奇偶分为两半，反应时的*Cronbach’sα*=0.54，自信度的*Cronbach’sα*=0.86，判断比例的*Cronbach’sα*=0.88。

检验因变量的效度，因为反应时曲线呈现倒U型，所以分别在比较刺激长度为40-50mm和50-60mm两个范围下计算因变量与比较刺激长度的Spearman相关系数。在比较刺激长度为40-50mm时，反应时与比较刺激长度显著相关，*rs*=0.55, *p*<.001；自信度与比较刺激长度显著相关，*rs*=0.88, *p*<.001；判断比例与比较刺激长度显著相关，*rs*=0.80, *p*<.001。在比较刺激长度为50-60mm时，反应时与比较刺激长度显著相关，*rs*=-0.54, *p*<.001；自信度与比较刺激长度显著相关，*rs=*0.86, *p*<.001；判断比例与比较刺激长度显著相关，*rs=*0.78, *p*<.001。

为检验因变量是否存在低限效应，我们选出两组自变量区间加以分析，第一组为46mm、48mm、50mm，第二组为50mm、52mm、54mm。

以第一组比较刺激长度为自变量，分别对被试的反应时、自信度和判断比例进行单因素重复测量方差分析。反应时的球形度检验不显著（*Mauchly’s W*=0.95, *p*=.182），比较刺激长度对反应时有显著性影响，*F*(2, 148)=46.73, *p*<.001, *partial* η2=0.39；使用Bonferroni法进行事后检验，46mm的反应时显著低于48mm的反应时，*MD*=-490.23, *p*<.001，46mm的反应时显著低于50mm的反应时，*MD*=-598.58, *p*<.001，48mm的反应时与50mm的反应时没有显著性差异，*MD*=-108.35, *p*=.249。自信度的球形度检验显著（*Mauchly’s W*=0.75, *p*<.001），使用Greenhouse-Geisser矫正结果，比较刺激长度对自信度有显著性影响，*F*(1.59, 102.10)=221.78, *p*<.001，*partial* η2=0.78；使用Bonferroni法进行事后检验，46mm的自信度显著低于48mm，*MD*=-29.69, *p*<.001，46mm的自信度显著低于50mm，*MD*=-57.95，*p*<.001，48mm的自信度显著低于50mm的自信度，*MD*=-28.26, *p*<.001。判断比例的球形度检验显著（*Mauchly’s W*=0.81, *p*<.001），使用Greenhouse-Geisser矫正结果，比较刺激长度对判断比例有显著性影响，*F*(1.68, 122.49)=261.04, *p*<.001，*partial* η2=0.78；使用Bonferroni法进行事后检验，46mm的判断比例显著低于48mm，*MD*=-14.73, *p*<.001，46mm的判断比例显著低于50mm，*MD*=-46.68, *p*<.001，48mm的判断比例显著低于50mm，*MD*=-31.95, *p*<.001。

以第二组比刺激长度为自变量，分别对被试的反应时、自信度和判断比例进行单因素重复测量方差分析。反应时的球形度检验显著(*Mauchly’s W*=0.69, *p*<.001)，使用Greenhouse-Geisser矫正结果，比较刺激长度对反应时有显著性影响，*F*(1.52, 108.15)=56.59, *p*<.001, *partial* η2=0.44；使用Bonferroni法进行事后检验，50mm的反应时显著高于52mm，*MD*=230.94, *p*<.001，50mm显著高于54mm，*MD*=619.98, *p*<.001，52mm的反应时显著高于54mm，*MD*=389.05, *p*<.001。自信度的球形度检验显著(*Mauchly’s W=*0.76, *p*<.001)，使用Greenhouse-Geisser矫正结果，比较刺激长度对自信度有显著性影响，*F*(1.61, 120.73)=243.67, *p*<.001, *partial* η2=0.77；使用Bonferroni法进行事后比较，50mm的自信度显著低于52mm，*MD*=-32.81, *p*<.001，50mm的自信度显著低于54mm，*MD*=-57.74, *p*<.001，52mm的自信度显著低于54mm，*MD*=-24.93, *p*<.001。判断比例的球形度检验显著(*Mauchly’s W*=0.67, *p*<.001)，使用Greenhouse-Geisser矫正结果，比较刺激长度对判断比例有显著性影响，*F*(1.50, 91.80)=380.23, p<.001, *partial* η2=0.86；使用Bonferroni法进行事后比较，50mm的判断比例显著低于52mm，*MD*=-38.80, *p*<0001，50mm的判断比例显著低于54mm，*MD*=-48.32, *p*<.001，52mm的判断比例显著低于54mm，*MD*=-9.52, *p*<.001。

再检验高限效应,我们选出两个自变量区间，第一组为40mm、42mm、44mm，第二组为56mm、58mm、60mm。

以第一组比较刺激长度为自变量，分别对被试的反应时、自信度和判断比例进行单因素重复测量方差分析。反应时的球形度检验显著*(Mauchly’s W*=0.83, *p*=.001)，使用Greenhouse-Geisser矫正结果，比较刺激长度对反应时有显著性影响，*F*(1.71, 124.92)=25.64, *p*<.001, *partial* η2=0.26；使用Bonferroni法进行事后比较，40mm的反应时显著低于42mm，*MD*=-58.92, *p*=.005，40mm的反应时显著低于44mm, *MD*=-164.06, *p*<.001，42mm的反应时显著低于44mm，*MD*=-105.14, *p*<.001。自信度的球形度检验显著(*Mauchly’s W*=0.80, *p*=.005)，使用Greenhouse-Geisser矫正结果，比较刺激长度对自信度有显著性影响，*F*(1.67, 80.00)=37.14, *p*<.001, *partial* η2=0.44；使用Bonferroni法进行事后比较，40mm的自信度显著低于42mm，*MD*=-2.84, *p*=.027，40mm的自信度显著低于44mm，*MD*=-9.58, *p*<.001，42mm的自信度显著低于44mm，*MD*=-6.74, *p*<.001。判断比例在三个条件下均为0%，因此比较刺激长度对判断比例没有显著性影响。

以第二组比较刺激长度为自变量，分别对被试的反应时、自信度和判断比例进行单因素重复测量方差分析。反应时的球形度检验显著(*Mauchly’s W*=0.60, *p*<.001)，使用Greenhouse-Geisser的矫正结果，比较刺激的长度对反应时有显著性影响，*F*(1.43, 99.88)=18.57, *p*<.001, *partial* η2=0.21；使用Bonferroni法进行事后比较，56mm的反应时显著高于58mm，*MD*=92.15, *p*=.001，56mm的反应时显著高于60mm，*MD*=124.64, *p*<.001，58mm的反应时显著高于60mm，*MD*=32.49, *p*=.043。自信度的球形度检验显著(*Mauchly’s W*=0.87, *p*=.027)，使用Greenhouse-Geisser矫正结果，比较刺激长度对自信度有显著性影响，*F*(1.77, 95.75)=36.29, *p*<.001, *partial* η2=0.40；使用Bonferroni法进行事后比较，56mm的自信度显著低于58mm, *MD*=-7.76, *p*<.001，56mm的自信度显著低于60mm, *MD*=-12.02, *p*<.001，58mm的自信度显著低于60mm, *MD*=-4.25, *p=*.003。判断比例在三个条件下均为100%，因此比较刺激长度对判断比例没有显著性影响。

4 分析与讨论

本实验结果显示，反应时的*Cronbach’sα*小于自信度和判断比例的*Cronbach’sα*，即反应时的信度不如自信度和判断比例，这与预期结果不符。反应时、自信度和判断比例的Spearman相关均显著，说明三者都有较高的效度。判断比例出现了低限效应和高限效应，反应时和自信度没有出现，说明反应时在敏感性上优于判断比例，但无法证明反应时在敏感性上优于自信度。

针对反应时的信度不如自信度和判断比例的结果，我们认为可能是速度—正确率权衡导致的。速度—正确率权衡是指在测验中，有些个体倾向于牺牲反应速度以提高正确率，而有些个体倾向于牺牲正确率以快速反应，当被试权衡的标准发生变化时，反应时也会相应变化（朱滢，2014）而且不同被试的权衡标准可能不同，这可能会影响反应时的分半信度。且Draheim等（2019）也总结了使用反应时作为因变量在关于个体差异和发展的研究中缺乏信度的问题。

反应时的敏感性并不高于自信度这一结果，与Petrusic和Baranski(2009)提出的主观反应时理论相符，他们指出实验中测量的反应时不一定能够反映个体客观的知觉—决策时间，即使在标准刺激与比较刺激差异明显的条件下，反应时仍然会受到自信度的影响，即自信度可视为比较刺激长度与反应时之间的中介变量，可能自信度本身有更高的敏感性。

综上所述，本实验结果显示反应时的信度低于自信度和判断比例；反应时、自信度、判断比例均有较好的效度；判断比例出现高限效应和低限效应，而反应时与自信度均未出现高/低限效应。本实验证明反应时优于判断比例，但未能证明反应时优于自信度。本实验中存在不足：未能很好地控制被试双眼到屏幕的距离，这对于结果可能有影响。有关反应时作为因变量的优越性还有待进一步研究。

**参 考 文 献**

Anstey, K. J., Lord, S. R., & Williams, P. (1997). Strength in the lower limbs, visual contrast sensitivity, and simple reaction time predict cognition in older women. *Psychology and Aging*, *12*(1), 137.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*(5), 508.

Eide, P., Kemp, A., Silberstein, R. B., Nathan, P. J., & Stough, C. (2002). Test-retest reliability of the emotional Stroop task: Examining the paradox of measurement change. *The Journal of Psychology*, *136*(5), 514-520.

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological review*, *18*(3), 186..

Hultsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(2), P101-P115.

Mislevy, R. J., Wingersky, M. S., Irvine, S. H., & Dann, P. L. (1991). Resolving mixtures of strategies in spatial visualization tasks. *British Journal of Mathematical and Statistical Psychology*, *44*(2), 265-288.

Petrusic, W. M., & Baranski, J. V. (2009). Probability assessment with response times and confidence in perception and knowledge. *Acta psychologica*, *130*(2), 103-114

Petrusic, W. M., & Baranski, J. V. (2009). Probability assessment with response times and confidence in perception and knowledge. *Acta psychologica*, *130*(2), 103-114.

Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & psychophysics*, *61*(1), 87-106.

Van Der Linden, W. J. (2009). Conceptual issues in response‐time modeling. *Journal of Educational Measurement*, *46*(3), 247-272.

Van Der Linden, W. J., Breithaupt, K., Chuah, S. C., & Zhang, Y. (2007). Detecting differential speededness in multistage testing. *Journal of Educational Measurement*, *44*(2), 117-130.

郭磊, 尚鹏丽, & 夏凌翔. (2017). 心理与教育测验中反应时模型应用的优势与举例. *心理科学进展*, *25*(4), 701.

赫葆源、张厚粲、陈舒永等编 实验心理学 北京大学出版社 1983年版 218-219

余柳涛, 鲍建樟, 陈清华, & 王大辉. (2016). 个体自信度对双人决策的影响. *心理学报*, *48*(8), 1013.

朱滢. (2014). 实验心理学 (第 3 版 ed.). 北京: 北京大学出版社.