Larsen, R. J., Diener, E., & Emmons, R. A. (1985). An evaluation of subjective well-being measures. *Social Indicators Research, 17*, 1–17.

Punch, K. F. (2005). *Introduction to social research. Quantitative and qualitative approaches*. London: Sage.

Segers, J. (1987). *Methoden voor de sociale wetenschappen*. Assen, The Netherlands: Van Gorcum.

Reliable Change Index

Martin Guhn[1], Barry Forer[2] and
Bruno D. Zumbo[3]
[1]Human Early Learning Partnership, School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[2]University of British Columbia, Vancouver, BC, Canada
[3]Department of ECPS & Department of Statistics, Institute of Applied Mathematics, University of British Columbia, Vancouver, BC, Canada

Synonyms

Critical difference; Clinical significance; Clinically significant change; Significant difference

Definition

Reliable Change Index (RCI) is a concept in measurement and assessment. An RCI is a psychometric criterion used to evaluate whether a change over time of an individual score (i.e., the difference score between two measurements in time) is considered statistically significant. Computationally, RCIs represent a ratio, in which the numerator represents an actual *observed difference score* between two measurements, and the denominator is some form of *standard error of measurement of the difference*. An RCI indicates whether an individual change score (e.g., between a patient's pre-intervention and post-intervention assessment) is statistically significantly greater than a difference that could have occurred due to random measurement error alone.

Description

The concept of *Reliable Change Index* (RCI) refers to a method that is used to test whether a change over time – that is, the difference score between two assessments of the same person at two points in time – may be considered "reliable" or "(clinically) significant." In particular, RCIs are commonly used to assess whether some condition or construct (e.g., depression; cognitive functioning) changed during an intervention (e.g., between pre-intervention and post-intervention).

The term "reliable change" is used to differentiate change that is reliable in the statistical sense (i.e., change that is statistically significant) from change that may have occurred due to random fluctuation in measurement (e.g. ▶ measurement error; Jacobson & Truax, 1991; Maassen, 2004).

Originally, Jacobson and Truax introduced an index to assess change in 1981, based on *Classical Test Theory*, building on previous work by McNemar (1962) and Lord and Novick (1968). The term *Reliable Change Index* was then introduced by Jacobson, Follette, and Revenstorf (1984). Today, the method is commonly referred to as the *Jacobson-Truax Index* or as the *classical approach* to reliable change. However, alternative methods to calculate RCIs have been developed, and the term RCI generally refers to a large number of different variations of reliable change indices that are based on similar concepts.

In essence, an RCI is designed to numerically quantify whether an observed difference between two measurements or assessments may be considered "reliable" or statistically significant. In principle, the formula for an RCI is simple: An RCI is calculated as a ratio, in which the numerator represents the difference between two measurements (e.g., a pre-intervention assessment of depression and a post-intervention assessment of depression) and the denominator represents some form of a *standard error of measurement of the difference*.

$$\text{Reliable Change Index} = x_{(\text{time 2})} - x_{(\text{time 1})}/$$

standard error of measurement of the difference

(1)

where $x_{(time\ 1)}$ and $x_{(time\ 2)}$ are measurement scores for an individual at different points in time.

In the original formulation of Classical Test Theory, the standard error of measurement of the difference was defined as the following (cf. McNemar, 1962):

$$\text{standard error of measurement of the difference} = \text{variance}_{x1} * (1 - r_{x1}) + \text{variance}_{x2} * (1 - r_{x2}) \tag{2}$$

where variance $_{x1}$ refers to the variance of all x scores at time 1, variance $_{x2}$ refers to the variance of all x scores at time 2, r_{x1} is the test reliability at time 1, and r_{x2} is the test reliability at time 2.

The formula shows that the standard error of measurement of the difference takes into account the variance of scores at time 1 and the variance of scores at time 2 as well as the reliability coefficients of time 1 and time 2. All methods for calculating the standard error of measurement of the difference are variations of this formula. The various methods differ in how they substitute – or estimate – the parameters of the formula (e.g., for cases in which the reliability coefficient and/or the variance of scores at time 2 is unknown, some methods assume/propose it is equal to the reliability coefficient and/or variance at time 1).

The size of an RCI is a direct estimate of the statistical significance of the difference score. Given that, in order to obtain an RCI, observed difference scores are divided by the corresponding standard error of measurement of the difference, RCI values are equivalent to standardized ▶ z-scores (i.e., they have a mean of 0 and a standard deviation of 1). Thus, an RCI that is greater than 1.96 denotes a statistically significant difference or, to use the current terminology, reflects a "reliable change." (A z-score of +1.96 corresponds to the 97.5th percentile of a normal distribution; in other words, 95 % of all z-standardized values in a normal distribution are smaller or equal to +/− 1.96. Following the convention of using 5 % as the threshold

for statistical significance, an RCI of greater than 1.96 is therefore considered statistically significant).

Despite the simplicity of the generic RCI formula, RCIs have been a matter of debate in the psychometric literature (Hinton-Bayre, 2000; Maassen, 2000; Mellenbergh & Van den Brink, 1998; Temkin, Heaton, Grant, & Dikmen, 1999). Primarily, the debate concerns the question of how to calculate the *standard error of measurement of the difference* – not to be confused with ▶ *standard error* or ▶ *standard error of measurement*, both of which are related, but different concepts. Due to different approaches to conceptualizing and calculating the standard error of measurement of the difference, numerous RCIs exist in the literature. In a review, Perdices (2005) presents eight different approaches to calculating the standard error of measurement of the difference. The different methods for calculating the standard error of measurement of the difference do not reflect trivial differences. In fact, Maassen (2004), Maassen, Bossema, and Brand (2009), and Perdices (2005) demonstrate how applying various RCIs to identical data may lead to a different standard error of measurement of the difference and, consequently, to different conclusions about the "significance" of a change score.

For a user of RCIs, it is critical to explicate the different statistical and conceptual assumptions that underlie the different methods for calculating RCIs. For example, the method proposed by Ley (1972) calculates the standard error of measurement of the difference by inserting the (generic) variance and retest reliability of a test (which, of course, need to be known, e.g., from a test manual/normative sample), as shown in formula (3).

$$\text{standard error of measurement of the difference} = \text{variance}_x * \text{sqrt}(2 * (1 - \text{retest reliability})) \tag{3}$$

In comparison, the formula by Christensen and Mendoza (1986) estimates its parameters from a sample under study, that is, by using the

variances of measurement scores at time 1 and time 2 and the covariance between measurement scores at time 1 and time 2, as illustrated by formula (4).

standard error of measurement of the difference
$$= \text{sqrt}(\text{variance}_{x1} + \text{variance}_{x2} - \text{vairance}_{x1} * \text{variance}_{x2} * \text{correlation}_{x1x2})$$

(4)

Except for in a few special scenarios, these two formulas lead to different results. Furthermore, the two methods are based on different assumptions and are designed for different scenarios. Ley uses a test's generic variance and test-retest reliability (given that they are known) to calculate the standard error of measurement of the difference. This approach assumes that a test's variance and its reliability are equal across samples and across measurement occasions.

Christensen and Mendoza' s approach (1986), on the other hand, takes into account that, for a given sample of interest, the variance of x scores may change between measurement time 1 and time 2, and it considers the covariance between the two measurement time points for the observed sample to be an adequate reference criterion against which to evaluate the significance of an individual change score.

Therefore, Ley's formula seems appropriate if an individual's change over time is to be evaluated in relation to a representative reference sample and if the retest reliability of the measure is known. Mendoza and Christiansen's method, on the other hand, can be used in a scenario in which measurements at time 1 and time 2 for a sample are available and in which individual change scores are to be evaluated in relation to the distribution of difference scores in the given sample. Choosing an appropriate RCI should therefore be guided by conceptual and theoretical considerations. However, as can be seen, in practice, the parameters of a given scenario limit which RCIs can possibly be calculated. As demonstrated above, some RCIs require that the test reliability be known; other RCIs require that the distribution of scores at time 1 and time 2 of a reference sample be known.

Conclusion and Further Resources

The discussion of the RCI presented here has focused on a generic description of the RCI as a concept in assessment. In addition, we briefly delineated some key challenges and limitations of RCIs that have been discussed in the psychometric literature. Given the substantially different conceptual and computational approaches of the various RCIs for calculating the standard error of measurement of the difference, it is critical that users indicate which RCI they have chosen to use, point out the RCI's potential limitations, and stringently explicate and evaluate the assumptions that underlie the chosen RCI.

Cross-References

▶ Measurement Error
▶ Reliability
▶ Significance, Statistical
▶ Standard Error of Measurement
▶ Standard Errors
▶ Standard Scores
▶ Test-Retest Reliability

References

Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*, 305–308.

Hinton-Bayre, A. (2000). Reliable change formula query. *Journal of the International Neuropsychological Society, 6*, 362–363.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336–352.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Clinical and Consulting Psychology, 59*, 12–19.

Ley, P. (1972). *Quantitative aspects of psychological assessment*. London: Gerald Duckworth.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Maassen, G. H. (2000). Principles of defining reliable change indices. *Journal of Clinical and Experimental*

Neuropsychology, 22, 622–632. doi:10.1076/1380-3395(200010)22:5;1-9;FT622.

Maassen, G. H. (2004). The standard error in the Jacobson and Truax reliable change index: The classical approach to the assessment of reliable change. *Journal of the International Neuropsychological Society, 10*, 888–893. doi:10.10170S1355617704106097.

Maassen, G. H. (2005). Reliable change assessment in sports concussion research: A comment on the proposal and reviews of Collie et al. *British Journal of Sports Medicine, 39*, 483–488. doi:10.1136/bjsm.2004.015594.

Maassen, G. H., Bossema, E., & Brand, N. (2009). Reliable change and practice effects: Outcomes of various indices compared. *Journal of Clinical and Experimental Neuropsychology, 31*, 339–352. doi:10.1080/13803390802169059.

McNemar, Q. (1962). *Psychological statistics* (3rd ed.). New York: Wiley.

Mellenbergh, G. J., & Van den Brink, W. P. (1998). The measurement of individual change. *Psychological Methods, 3*, 470–485.

Perdices, M. (2005). How do you know whether your patient is getting better (or worse)? A user's guide. *Brain Impairment, 6*, 219–226. doi:10.1375/brim.2005.6.3.219.

Temkin, N. R., Heaton, R. K., Grant, I., & Dikmen, S. S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society, 5*, 357–369.

Religion

Religion and Coping with Child's Death

Religion and Politics in the Middle East

Religion and Sexism

Lauren E. Maltby[1] and Elizabeth L. Hall[2]
[1]Harbor-UCLA Medical Center, Torrance, CA, USA
[2]Biola University, Rosemead School of Psychology, La Mirada, CA, USA

Synonyms

Ambivalent sexism; Bigotry and religion; Chauvinism and religion; Gender harassment; Sexist attitudes; Sexist behaviors; Sexist beliefs

Definition

From a social psychological perspective, sexism includes three different, though related, concepts: sexist attitudes (hostile and benevolent), sexist beliefs (gender stereotypes and gender ideologies), and sexist behaviors (discrimination and its subtype, harassment). In this model, sexist attitudes and beliefs lead to sexist behaviors, or discrimination. Ambivalent sexism theory, introduced by Glick and Fiske in 1996, emphasizes that sexism must be understood within a social context in which men and women are interdependent, that is, they are intimately brought together in the context of heterosexual romance, domestic life, and child rearing. Consequently, rather than being characterized by an attitude of hostility, as is common in most forms of prejudice, sexism is characterized by ambivalence, comprised of both hostility and benevolence. These hostile and benevolent attitudes are closely tied to gender stereotypes and gender ideologies. Gender stereotypes attribute certain characteristics to women and men, while gender ideologies are more general prescriptive beliefs that include stereotypes as well as broader beliefs about proper roles for men and women in society.

The presence of sexist attitudes and beliefs may lead to discrimination. Crosby and Stockdale (2007) state that sex discrimination occurs "when a person is or people are treated unfairly... because of gender" (p. 3). It can be seen in unfair levels of