

JoVE: Science Education

Reliability

--Manuscript Draft--

Manuscript Number:	10046
Full Title:	Reliability
Article Type:	Manuscript
Section/Category:	Manuscript Submission
Corresponding Author:	Gary Lewandowski UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Gary Lewandowski
First Author Secondary Information:	
Order of Authors:	Gary Lewandowski David Strohmets Natalie Ciarocco
Order of Authors Secondary Information:	

PI: Gary Lewandowski, David Strohmetz & Natalie Ciarocco

Psychology Education Title:

Reliability

Overview: In order to study something scientifically, a researcher needs to determine a way to quantify it. However, psychological constructs can be challenging to measure and quantify. This video examines reliability in the context of content analysis.

A recent study in the journal *Pediatrics* reported that 4-year-olds who watched a fast-paced cartoon (i.e., *SpongeBob SquarePants*) had worse performance on cognitive tasks such as following rules in a game, listening to direction from an adult, and delaying gratification compared to other children who watched a slower paced cartoon. In addition to pace of the cartoon, the content of the cartoon may also have deleterious effects on its young viewers.

This video uses a simple two-group design, to exemplify the issue of reliability, in examining the question of whether the cartoon *SpongeBob SquarePants* has more “inappropriate content” than does the cartoon *Caillou*.

Procedure:

1. Define Key Variable

- 1.1. Create an operational definition (i.e. a clear description of exactly what a researcher means by a concept) of “inappropriate content.”
- 1.2. Consult definitions created by the organization TV Parental Guidelines and approved by the Federal Communications Commission.
 - 1.2.1. “Inappropriate content” is any crude or rude behavior (e.g., toilet humor), depictions of verbal or physical aggression (e.g., name calling, hitting, etc.), bad language (e.g., curse words), or references to drug use, violence, or sex.

2. Create coding categories from the operational definition of inappropriate content.
 - 2.1. We need a systematic process for the researcher participants (referred to here as the “Raters”) to identify instances of the targeted behavior, inappropriate behavior, so we create this log (table 1).
3. Raters separately watch the same episode of SpongeBob SquarePants and provide codings.
4. Raters separately watch the same episode of Caillou and provide codings.
5. We compare ratings to see if the Raters came up with similar ratings for each show.
 - 5.1. “Reliability” is the ability to consistently measure the variable – here inappropriate content.
 - 5.2. “Inter-rater reliability” is the ability for more than one person to measure the variables and for their measurements to be in accord.

Results:

Table 2

Figure 1.

Figure 2.

The results indicate that the Raters had a high level of agreement or consistency in their ratings within each cartoon episode, which indicates high inter-rater reliability.

There is also reliability or consistency in SpongeBob SquarePants episodes having more inappropriate content than Caillou.

Also, we start to see how some raters may have individual biases. For example Rater 3 tends to see more inappropriate content in SpongeBob than the other 2 raters and Rater 1 tends to see less in Caillou than other raters.

Comment [AK1]: We have a result that demonstrates inter-rater reliability, but what about reliability? One might show that sponge bob episodes are consistently rated more inappropriately by the same scorer when compared with different Caillou episodes.

Applications:

Researchers have increasingly turned their attention toward analyzing television's content, especially as it relates to children. As discussed prior to this current experiment, a recent study in the journal *Pediatrics* correlates the fast pace of the SpongeBob SquarePants cartoon to relatively poor cognitive abilities in the children who watch it.

Since the results of our experiment appear reliable, future research could examine whether the relative amount of inappropriate content in SpongeBob is also (or alternatively) responsible for children's lower cognitive performance after watching.

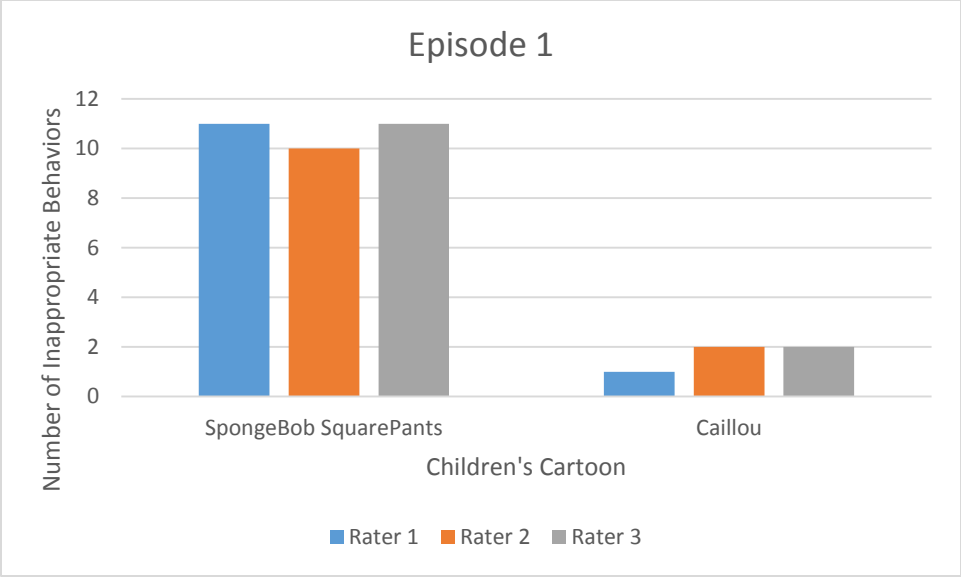
One of the most important applications of reliability is in the use of survey instruments. Researchers must be sure that participants will consistently answer each of the items in a particular scale. That is, in a 5 item measure of life satisfaction, participants should answer items 1 & 2 in a somewhat similar fashion to how they answer questions 3, 4, and 5. In addition, researchers want to make sure that their measurement in an experiment are consistent over time. So if a researcher is using pupil dilation to indicate interest in a stimulus, the researcher must be sure that pupil dilation is a consistent indicator of interest.

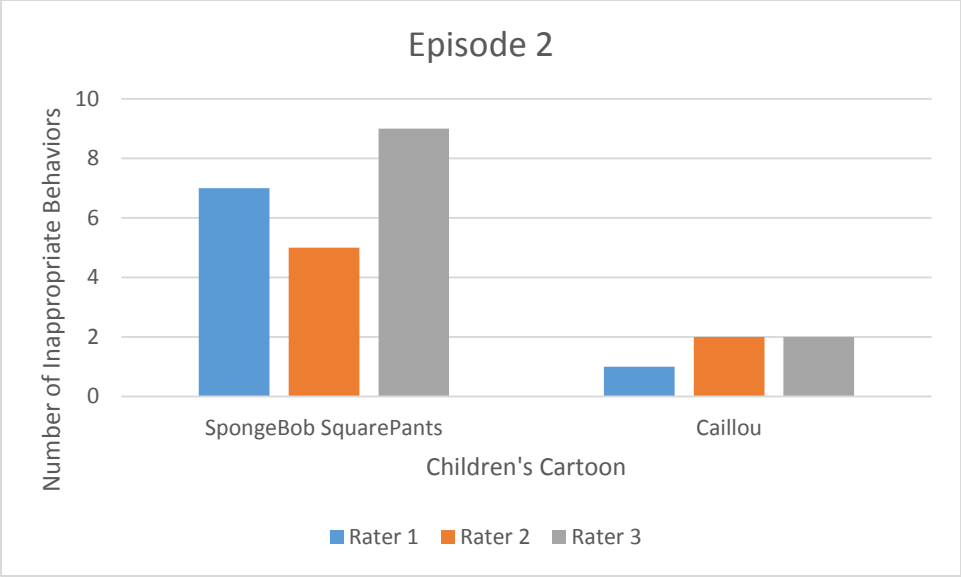
Comment [AK2]: The application content should explain how the main topic is applied. The video is about reliability in the quantification of psychological constructs. The applications should discuss how reliability is important for other forms of experiments.

Legend:

Figure 1. Instances of Inappropriate Content by Rater and Cartoon for Episode 1

Figure 2. Instances of Inappropriate Content by Rater and Cartoon for Episode 2





<u>Coding Categories</u>	<u>Themes and Exemplars</u>	<u>Count</u>
Crude Behavior	Toilet humor Purposefully disgusting behaviors	
Rude Behavior	Disrupting others Poor Manners	
Language	Using curse words	
Verbal Aggression	Insults Yelling Name-Calling	
Physical Aggression	Hitting Pushing/Shoving Tripping	
Drug References	Verbal (suggestive statements/conversation) Nonverbal (mimicking drug use)	
Sexual References	Verbal (suggestive statements/conversation) Nonverbal (mimicking sexual acts)	

		<u>SpongeBob</u>			<u>Caillou</u>		
<u>Coding Categories</u>	<u>Themes and Exemplars</u>	<u>Rater 1 Count</u>	<u>Rater 2 Count</u>	<u>Rater 3 Count</u>	<u>Rater 1 Count</u>	<u>Rater 2 Count</u>	<u>Rater 3 Count</u>
Crude Behavior	Toilet humor Purposefully disgusting behaviors						
Rude Behavior	Disrupting others Poor Manners						
Language	Using curse words						
Verbal Aggression	Insults Yelling Name-Calling						
Physical Aggression	Hitting Pushing/Shoving Tripping						
Drug References	Verbal (suggestive statements/conversation) Nonverbal (mimicking drug use)						
Sexual References	Verbal (suggestive statements/conversation) Nonverbal (mimicking sexual acts)						