Learning and Individual Differences 32 (2014) 19–25

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Learning to judge creativity: The underlying mechanisms in creativity training for non-expert judges

Martin Storme [a,b,*], Nils Myszkowski [a,b], Pinar Çelik [a], Todd Lubart [a]

[a] Laboratoire Adaptations Travail Individu, Université Paris Descartes, 71 Avenue Edouard Vaillant, Boulogne-Billancourt, France
[b] Ecole Supérieure du Commerce Extérieur, 10 rue Sextius Michel, Paris, France

ARTICLE INFO

Article history:
Received 25 February 2013
Received in revised form 27 January 2014
Accepted 9 March 2014
Available online xxxx

Keywords:
Creativity judgment
Training
Learning mechanism

ABSTRACT

Evaluating individual creativity is an important challenge in creativity research. We developed a training module for non-expert judges in which participants learned the definitions of components of creativity and received expert feedback in an interactive creativity judgment exercise. We aimed to test whether and how the training module would increase the reliability and validity of non-expert ratings. Study 1 (N = 79) showed that the training had a positive effect on the test–retest reliability and validity of creativity ratings. Study 2 (N = 126) replicated the results on test–retest reliability and validity but with low absolute values, indicating that trained participants cannot substitute experts. In addition, Study 2 showed that the effect of the training module on the validity of creativity ratings was mediated by increased validity of ratings of novelty and elaboration. The results are discussed in terms of theoretical and practical relevance.

1. Introduction

Despite a growing interest in creativity research, the field is often challenged because of the difficulty in finding reliable and valid methods to assess individual creativity. In the literature various measures of individual creativity, including divergent thinking tests, attitude inventories, ratings by peers, and judgments of products, are described (Lubart, 1994). One of the most used techniques to evaluate individual creativity is the Consensual Assessment Technique (Amabile, 1982, 1996; Hennessey, 1994; Kaufman, Baer, & Cole, 2009; Kaufman, Baer, Cole, & Sexton, 2008; Kaufman, Gentile, & Baer, 2005; Runco, 1989). This technique consists of assessing an individual's level of creativity by assessing the creativity level of certain products made by this individual. The technique is based on giving minimal information about creativity to a group of judges who have to rate the creativity of a set products relative to other products in the sample (Amabile, 1996; Dollinger & Shafran, 2005). Because creativity is a relative construct it is commonly expected that judges should be familiar with the domain, i.e. experts or, at least, gifted novices (Amabile, 1996; Dollinger & Shafran, 2005; Kaufman et al., 2009). Several studies show that experts have a good interrater agreement (Kaufman et al., 2009; Kaufman et al., 2008; Kaufman, Baer, Cropley, Reiter-Palmon, & Sinnett, 2013; Kaufman et al., 2005). A potential difficulty, however, in using this technique is finding and asking experts to evaluate sometimes hundreds of products; this may pose a real challenge to creativity researchers (Cropley & Kaufman, 2012). The aim of the present paper is twofold. From a practical viewpoint we aim to investigate whether it is possible to teach novices how to evaluate the creativity of products made by individuals. From a more conceptual/theoretical viewpoint a parallel aim that we have is to shed more light on the underlying processes of learning to judge creativity. By doing this we aim to understand how exactly people arrive at creativity judgments.

1.1. Previous research

Recently, Dollinger and Shafran (2005) suggested that non-expert judges could become more familiar with a domain of products by exposing them their prototypes. They conducted an experiment in which they compared expert judges' creativity ratings of drawings (made by university students) with trained non-expert judges' ratings of the same drawings. The training module consisted of presenting 16 representative drawings to non-expert judges before they had to evaluate the target drawings. Their results showed that trained judges' and expert judges' ratings loaded on a single principal component; thus expert and trained non-expert judges' mean creativity ratings were highly correlated. They concluded that showing non-expert judges representative drawings previous to the actual creativity judgment can make their creativity ratings more similar to expert judges' creativity ratings. However, because this study had no control group, it is difficult to assess causality and the extent to which their training module increased the validity of non-expert judges' creativity ratings compared to a baseline. Considering the number of participants in their study – 5 expert judges and 5 non-

* Corresponding author at: Institut de Psychologie, 71 Avenue Edouard Vaillant, Boulogne-Billancourt, France.
E-mail address: martinstorme@ymail.com (M. Storme).

http://dx.doi.org/10.1016/j.lindif.2014.03.002
1041-6080/

expert judges – it is even more critical to have a baseline to which they could have compared the trained group. Dollinger and Shafran (2005) suggested that further research should be conducted to generalize their results regarding the possibility to enhance the expertise of non-expert judges with a training module.

According to most creativity researchers (Besemer & O'Quin, 1999; Caroff & Besancon, 2008; Runco & Charles, 1993; Storme & Lubart, 2012) creativity ratings of a product typically involve rating subcomponents of creativity, such as novelty, resolution and elaboration. These subcomponents are used in various scales such as the Creative Product Semantic Scale (CPSS; Besemer & O'Quin, 1999; White & Smith, 2001; White, Shen, & Smith, 2002; O'Quin & Besemer, 2006) and the Creative Solution Diagnosis Scale (CSDS; Cropley & Kaufman, 2012), to judge creative ideas, products or designs. A study conducted by Cropley and Kaufman (2012) suggests that novices can be reliable judges when they are provided with precise criteria to assess creativity. The study aimed to develop the Creative Solution Diagnosis Scale, based on four subcomponents of creativity (relevance, novelty, elegance and genesis). The results showed that novices had a high level of interrater agreement (Cronbach's α ranging between .87 and .98) when rating the same products. Note that in this study the researchers did not assess the validity of the novices' creativity ratings by comparing their ratings with ratings of experts, but nonetheless the study shows that relevance, novelty, elegance and genesis are meaningfully related to each other. It appeared that these four components of creativity loaded on a single factor which the authors called 'functional creativity'.

In sum, previous research suggests that novices may learn to judge creativity as long as they have knowledge of the existing products in the domain to be rated (Dollinger & Shafran, 2005) and knowledge of the subcomponents of creativity (Cropley & Kaufman, 2012). Yet to our knowledge there are no existing studies that directly investigated the role of these subcomponents when novices are taught to judge creativity. In the present paper, we report on two studies in which we implement a new training module that we compare to a control group of individuals who are not trained. To give our trainees an analytical understanding of what creativity is, we provide them with a precise definition of creativity involving the subcomponents of creativity. This allows us to investigate the underlying mechanisms involved in the learning of rating creativity. In addition, our trainees are not only passively exposed to representative prototypes of the products to be judged, as in Dollinger and Shafran's (2005) study; but they also receive expert feedback in an interactive creativity judgment exercise.

Our aim was twofold. First, we investigated whether our training module, compared to a control module, would increase the reliability and validity of creativity judgments. Second, to show the underlying mechanism of learning to judge creativity, we investigated whether the effect of the training could be (partially) explained by the extent that trainees made correct use of the subcomponents of creativity.

1.2. Principles of the training module

In this section we describe the basic principles behind our training module. We will elaborate further on the procedural details of the training module in the Method section. In order to study the effect of training non-expert judges, we developed a specific training module based on children's drawings. The training module consists of two stages. The first stage consists in providing participants with a definition of creativity based on explaining to them that products (in this case drawings) differ in creativity because they differ in elaboration, novelty and resolution. More specifically, they are explained that they should take these three subcomponents into account when deciding on the creativity level of a given drawing. In this first stage trainees are also shown prototypical drawings to familiarize them with different levels of creativity. Thus, the first stage aims to give participants an analytical understanding of creativity and familiarize them with representative drawings of varying levels of creativity.

In the second and last stage participants get an opportunity to exercise judging creativity levels of a new set of drawings. First, they are asked to rate the creativity level of a set of drawings. Then they receive feedback on how well they did. Compared with the first stage, the second stage is a more interactive feedback stage in which participants learn to rate creativity.

To sum up, the training consisted of first presenting participants with a precise definition of creativity, then showing participants prototypical drawings, and finally providing participants with feedback on the accuracy of their creativity judgments in a creativity rating exercise. We expected that the training, compared to a control condition, would improve the reliability and validity of creativity judgments in a final set of drawings, and that this would be mediated by improvements in the validity of judgments of the subcomponents of creativity.

1.3. Overview of the studies

This article reports on the results of two studies which aimed at testing the effectiveness of a training module for judging creativity, intended for non-expert individuals, and to investigate learning mechanisms of creativity judgments. Our first study aimed at investigating whether it is possible to learn to judge creativity, and the size of the effectiveness of our training module. We hypothesized that trained participants would produce more reliable and more valid creativity ratings than non-trained participants. To this end, we assessed whether trained participants agreed more with each other (i.e. had higher interrater reliability) than non-trained participants. We also assessed whether trained participants were more stable over time (i.e. had higher temporal stability) in their creativity judgments than non-trained participants. Finally, we assessed whether the ratings of trained participants, compared to the ratings of non-trained participants, were more reliable and more in agreement with the ratings of expert judges (i.e. had higher validity).

The second study aimed at investigating the underlying psychological mechanisms of learning to judge creativity. Because the training module provided our participants with a precise definition of creativity based on novelty, elaboration and resolution, we expected that the training module would also improve the trained participants' validity of novelty and elaboration judgments. More specifically, we expected that the more valid the judgments of novelty and elaboration are (i.e. the more in agreement with expert judges' evaluations of novelty and elaboration), the more valid the judgments of creativity would be as well. In other words, we hypothesized that increased validity of novelty and elaboration judgments would mediate the relationship between the training module and the validity of creativity judgments.[1]

2. Study 1: global effect of the training on reliability and validity

Participants were randomly assigned to a training or control module. The control module was comparable to the training module to the extent that participants were exposed to the same drawings as in the training module, but participants received no definition of creativity and no feedback on their accuracy in the exercise session. Including this condition to our study design allowed us to rule out alternative explanations for the effect of the training, and assess the extent to which the training had an effect compared to a baseline of non-trained participants.

[1] We decided to limit the number of dimensions to be evaluated by the participants to make the judgment process not too complex. Since the products to be evaluated were children's drawings, we dropped judgments of resolution/usefulness from the ratings.