British Journal of Psychology (2010), 101, 47–68
© 2010 The British Psychological Society

The British Psychological Society

www.bpsjournals.co.uk

The selection of creative ideas after individual idea generation: Choosing between creativity and impact

Eric F. Rietzschel[1]*, Bernard A. Nijstad[2] and Wolfgang Stroebe[3,4]

[1]Department of Social and Organizational Psychology, University of Groningen, Groningen, The Netherlands
[2]Department of Work and Organizational Psychology, University of Amsterdam, Amsterdam, The Netherlands
[3]Department of Social and Organizational Psychology, Utrecht University, Utrecht, The Netherlands
[4]CREMA – Center for Research in Economics, Management and the Arts, Switzerland

It is commonly assumed that successful innovation depends on creative idea generation: the more ideas are generated, the higher the probability of selecting a very good idea should be. However, research has shown that people do not perform optimally at idea selection and that ideational output may not contribute much to creative idea selection. The present studies aim to explain this phenomenon. We identified the strong tendency of our participants to select feasible and desirable ideas, at the cost of originality, as the main reason for their poor selection performance. Two manipulations of participants' processing of the available ideas (exclusion instructions and quality ratings) had no effect on selection effectiveness. In contrast, explicitly instructing participants to select creative or original ideas did improve selection effectiveness with regard to idea originality, but at the same time decreased participants' satisfaction and the rated effectiveness of chosen ideas. Results are discussed in relation to an effectiveness–originality trade off.

In all domains of society, progress depends on the adoption of new procedures or products. Such innovation necessarily starts with the generation of creative ideas. It is commonly assumed that facilitation of this idea generation stage is crucial, and much research has addressed the factors that stimulate or inhibit idea generation (see Diehl & Stroebe, 1987; Mullen, Johnson, & Salas, 1991; Paulus & Brown, 2003, for overviews). The underlying assumption is that it is vitally important to generate as many creative

* *Correspondence should be addressed to Dr Eric F. Rietzschel, Department of Social and Organizational Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands (e-mail: e.f.rietzschel@rug.nl).*

DOI:10.1348/000712609X414204

ideas as possible, because this will increase the probability that at least one of these ideas is extremely good ('quantity leads to quality'; Osborn, 1953). For this reason, many creativity techniques, such as the well-known brainstorming technique (Osborn, 1953), are specifically designed to maximize participants' ideational output.

However, idea generation is only part of the innovative process, and the availability of creative ideas is a necessary, but insufficient condition for innovation (e.g. Nijstad & De Dreu, 2002; West, 2002). For actual implementation of creative ideas, the best ideas must be recognized and selected from the pool of generated ideas. For a long time, creativity researchers assumed that participants in idea generation sessions (such as brainstorming) would also be able to identify the most creative idea. However, upon testing this hypothesis, it was found that people perform very poorly at selecting creative ideas (Faure, 2004; Putman & Paulus, in press; Rietzschel, Nijstad, & Stroebe, 2006). So far, this striking phenomenon has not been explained.

In this article, we develop and test several potential explanations. We first review previous research on idea selection, and then discuss two factors that might improve selection effectiveness: (1) the degree to which people actually consider all available alternatives, and (2) the nature of the explicit and implicit selection criteria that people use. We then report two studies in which the role of these factors in idea selection is examined.

The selection of high quality ideas

In the literature on creative idea generation, idea quality is usually defined as a combination of originality and feasibility (e.g. Diehl & Stroebe, 1987). This fits with the approach generally taken by creativity researchers (e.g. Amabile, 1996; Runco & Charles, 1993; Sternberg & Lubart, 1999). Originality is the hallmark of creative behaviour, and ideas will not be creative if they are not new or unusual. However, some form of appropriateness is essential to distinguish truly creative behaviour from behaviour that is merely bizarre or erratic. Likewise, although ideas must be original in order to be called creative, they will not be implemented if they are not feasible. Hence, the usual definition of a 'good' idea is an idea that is both highly original (or unusual) and highly feasible (or useful).

Thus far, three studies have addressed the combination of creative idea generation and idea selection. Putman and Paulus (in press) studied the effect of group versus individual brainstorming on selection performance. Whereas idea generation took place in interactive or nominal groups (in a nominal group, people generate ideas individually, but their output is pooled), idea selection was conducted in interactive groups (with members of nominal groups selecting ideas from their pooled production). In line with previous research (see Diehl & Stroebe, 1987; Mullen *et al.*, 1991), Putman and Paulus found that nominal groups generated more ideas than interactive groups. They also found that the ideas generated by nominal groups were, on average, more original than those generated by interactive groups. Most importantly, however, the average originality of the *selected* ideas was higher for groups who had generated ideas as nominal groups than for groups who had generated ideas as interactive groups. However, both individuals and groups rarely selected their best ideas (as assessed by independent raters).

Faure (2004) conducted a similar study, but with slightly different results. Participants in this study first generated ideas, either as an interactive group or individually as a nominal group. Next, every participant made an individual pre-selection

of his or her five favourite ideas from the group's total production. Finally, the group members together selected the three best ideas from the pre-selected ideas. Faure also found that nominal groups generated more ideas, and more original ideas, than interactive groups. However, she found no difference between interactive and nominal groups with regard to the quality of selected ideas. In other words, the higher productivity of nominal groups, and the resulting higher availability of original ideas, was not associated with the selection of better ideas.

Rietzschel *et al.* (2006) also reported a study in which participants generated and selected ideas. Both idea generation and idea selection took place in interactive groups or in nominal groups. Thus, unlike in the studies by Putman and Paulus (in press) and Faure (2004), members of nominal groups did not see each other's ideas. The ideas generated by nominal groups were more original, and less feasible, than those generated by interactive groups. Moreover, nominal groups generated a larger number of good ideas. However, these differences were not present for the average originality and feasibility of the selected ideas. Thus, consistent with the findings of Faure (2004), the higher availability of good ideas in nominal groups did not cause these individuals to select more creative ideas than groups. Moreover, neither originality nor feasibility of selected ideas were significantly correlated with productivity. Most striking was the finding that the average quality of selected ideas was not higher than the average quality of the generated ideas. Thus, participants' idea selection was in fact not better than chance.

To summarize, only Putman and Paulus (in press) found a significant difference between interactive and nominal groups on idea selection, but this effect was not obtained by Faure (2004) and Rietzschel *et al.* (2006). The other results are highly consistent across the three studies and suggest two important conclusions: (1) a higher productivity after a brainstorming session is associated with higher availability, but not necessarily selection, of creative ideas, and (2) selection effectiveness in these studies was not optimal: although good ideas were available, this potential was not (fully) realized in the selection stage. Because it is clear that an ineffective selection procedure can easily lead to suboptimal results even when a prior idea generation session was highly effective, we decided to focus on increasing selection effectiveness. We hypothesize that there are two particularly likely candidates for poor idea selection performance: a lack of clear selection criteria and shallow processing of the available ideas. It follows that providing clear criteria and improving processing of ideas will lead to better idea selection.

Criteria and processing

A striking feature of the idea selection studies discussed above is that participants in these studies were simply instructed to select the 'best' ideas, or their 'favourite' ideas, but that ideas were evaluated in terms of originality and feasibility, or a combination of the two. It seems plausible that the efficacy of participants' selection could be improved, if they are given clear instructions about the criteria according to which they should select ideas. Furthermore, if lack of clearly defined selection criteria leads to poor selection performance, one wonders which criteria people spontaneously apply to idea selection. Although in the creativity literature idea quality is commonly assessed with the dimensions originality and feasibility, this does not mean that 'lay' participants will also use these criteria when making their selection. Given that creativity is often considered to be a very vague concept (Sternberg & Lubart, 1999), it is possible that

participants who are not provided with specific selection criteria simply have no idea what exactly is expected of them, and can therefore hardly be expected to make an effective selection with regard to the quality dimensions the researchers happen to use.

Another problem with creative idea selection is that people typically have to make their selection from a very large number of alternatives. For example, a brainstorming session can easily yield dozens of ideas. In general, people who have to choose from many alternatives tend to make use of simplifying strategies in order to reduce the cognitive complexity of the decision (e.g. Simon, 1955; Tversky, 1972; see Hastie & Dawes, 2001, for an overview). Even if their goal is to select one single best option, people often engage in pre-choice screening of the available alternatives (e.g. Beach, 1993), so that more cognitive resources become available for careful consideration of these remaining options (see also Parks & Cowlin, 1995). Because people who want to select the best ideas from their own production will have to choose from a large number of alternatives, a simplifying strategy is likely to be used. This may even undo the potentially beneficial effect of explicit selection criteria. If so, selection criteria will only improve selection performance if participants are simultaneously induced to pay more attention to the available alternatives. This could be attained by having participants make a more careful pre-selection.

STUDY 1

To address the issue of selection criteria and systematic processing, we conducted a study in which participants first generated ideas, and then selected their best idea. The presence of selection criteria was manipulated by instructing half of the participants to select an idea that was original and feasible. The other participants did not receive specific selection criteria, but were instructed to select 'the best' idea. Assuming that participants are able to recognize original and feasible ideas, it was hypothesized that explicit selection criteria would improve the effectiveness of the selection process, and would therefore lead to the selection of better (i.e. more original and feasible) ideas.

We further gave participants specific instructions for their pre-selection. When making a pre-selection from the available alternatives, people can either pre-select the options from which they will make their final choice (positive), or eliminate the options which are unacceptable (negative). Although these two strategies are logically equivalent, people do not always treat them as such. For example, Shafir (1993) found that people treat available choice options differently, depending on whether they are instructed to *choose* or to *reject* an option. Following up on this finding, Yaniv and Schul (1997) found that participants who were instructed to use a negative (i.e. exclusion) pre-selection strategy formed a larger consideration set than participants who used a positive (i.e. inclusion) strategy. Furthermore, participants with exclusion instructions more often chose the correct answer, presumably because a larger consideration set was more likely to contain the correct answer. This was confirmed in a study by Heller, Levin, and Goransson (2002).

If the poor selection performance found in earlier studies was due to participants' prematurely discarding good ideas, this problem might be avoided by giving participants exclusion instructions. Explicit instructions to form a consideration set should lead participants to consider each individual idea, thus increasing the chance that good ideas get selected. Furthermore, the larger consideration set formed under exclusion instructions is more likely to contain the best ideas, simply because it contains more

ideas; this could lead to an even stronger increase in selection effectiveness. Hence, we had our participants use an inclusion or exclusion strategy; we also employed a control condition where participants received neither inclusion nor exclusion instructions.

We expected that participants with exclusion instructions would form a larger consideration set than participants with inclusion instructions. We further expected that selection criteria would lead to a more effective selection process, and that this in turn would lead to the selection of better ideas. Further, we expected the latter effect to be most pronounced in combination with exclusion instructions.

Method

Participants

Fifty-five undergraduate psychology students (45 women, 10 men, mean age = 20.1 years) participated in this experiment for course credit or 7 euros (about 9 US dollars).

Independent variables

The experiment had a 3 (selection procedure: inclusion, exclusion, or direct selection) × 2 (instructions: criteria or no criteria) factorial design. Participants were assigned randomly to one of the experimental conditions. All participants worked individually throughout the experiment.

Procedure

The procedure consisted of the following tasks: idea generation (15 min), pre-selection (no time limit), final selection (no time limit), and exit questionnaire (no time limit). The pre-selection task was performed only by participants in the inclusion and exclusion conditions; participants in the direct selection condition made their final selection immediately after the idea generation task.

Upon arrival in the laboratory, participants received and read the first instructions (which were identical for all conditions). It was explained that they were about to participate in a brainstorming session, and that they would generate ideas about possible improvements in the education at the Department of Psychology. At this point, the four brainstorming rules (quantity is wanted, freewheeling is welcomed, combine and improve ideas, and criticism is ruled out) were explained. Participants were advised to keep these rules in mind during idea generation and were instructed to be concise and to the point in the phrasing of their ideas. After idea generation, participants in the inclusion condition and the exclusion condition received the instructions for the pre-selection task, whereas participants in the direct selection condition received the instructions for the final selection task.

In the pre-selection task, participants were asked to read through the ideas they had just generated and to make a pre-selection from these ideas. Participants in the inclusion condition were instructed to mark those ideas that they thought were good enough for further consideration; participants in the exclusion condition were instructed to cross out those ideas that they thought were *not* good enough for further consideration. Half of the participants (criteria condition) were told that 'a good idea is both *original* (innovative and unusual) and *practically feasible*' and were instructed to use these criteria in making their pre-selection. The other participants did not receive specific criteria upon which to base their selection (no criteria condition); these participants

were simply instructed to mark (cross out) those ideas that were 'good enough' (not good enough) for further consideration.

The next task consisted of making the final selection. Participants in the inclusion condition were instructed to select the best idea from those ideas that they had marked in the pre-selection task; participants in the exclusion condition were instructed to select the best idea from those ideas that remained after the pre-selection task; participants in the direct selection condition were instructed to select the best idea from the ideas they had generated in the idea generation task. Participants in the criteria condition were given the same selection criteria as earlier, and were instructed to select the idea that best satisfied these criteria; the other participants were simply instructed to select 'the best idea'.

After making the final selection, participants completed the post-experimental questionnaire. After completing this questionnaire, participants were debriefed, thanked and paid.

Measures and dependent variables

Productivity

For each participant, all unique generated ideas were counted to yield a measure of productivity.

Consideration set

In the inclusion condition, the consideration set was defined as the number of ideas marked in the pre-selection task. In the exclusion condition, the consideration set was defined as the number of ideas retained (i.e. not crossed out) in the pre-selection task. Because participants in the direct selection condition went immediately to the final selection, no consideration set was computed for these participants.

Idea quality

A trained rater who was blind to conditions coded all ideas for originality and feasibility (on a five-point scale; 1 = 'not at all original/feasible', 5 = 'extremely original/feasible'). Training consisted of elaborate explanation of the concepts of originality and feasibility. Moreover, on the rating scales, the scale points were accompanied by descriptions. For example, unoriginal ideas (with an originality of 1) were explained to be very common, often concerning issues that already exist (rather than being new). In contrast, highly original ideas (with an originality of 5) were described as ideas that are mentioned rarely, are very innovative, and often introduce radically new applications of existing things or things that are completely new. Similarly, the explanations for feasibility referred to practical issues; whereas unfeasible ideas cannot be implemented because the means are nonexistent or unavailable, highly feasible ideas are very easy to implement and would not require large investments, either in money or in time. To further safeguard rating validity, all ideas were first rated in random order (i.e. randomized across the entire idea pool) for originality, and then in a newly randomized order for feasibility. The first author coded a random subset of 250 ideas, which constituted about 30% of the total pool of generated ideas. As a measure of inter-rater agreement, we computed the intra-class correlations, using a two-way random model and consistency definition (e.g. McGraw & Wong, 1996; Shrout & Fleiss, 1979); these were .74 for originality and

.67 for feasibility. These values are considered 'good' to 'excellent' according to criteria specified by Cichetti and Sparrow (1981). Furthermore, with raters considered to be in agreement when their ratings differed by no more than one point (Diehl & Stroebe, 1987), agreement existed in 97.6% of the cases for originality and 99.6% of the cases for feasibility.

We then computed two originality measures for each participant: the average originality of the generated ideas and the originality of the idea selected in the final selection task. The same procedure was followed with the feasibility scores. Moreover, for each participant we averaged the originality and the feasibility of the selected ideas into an aggregate quality measure.

Questionnaire items

All items in the questionnaire were rated on five-point scales (1 = 'not at all', 5 = 'very much'). Four items asked participants whether they felt compelled to use originality and feasibility as selection criteria, and whether they had found it important to use originality and feasibility as selection criteria. Two other items measured whether participants were satisfied with the selection process, and whether they felt certain that they had selected the best idea.

Results

Selection criteria

We analysed the questionnaire items concerning selection criteria with 3 (selection procedure: inclusion, exclusion, or control) × 2 (instructions: criteria or no criteria) ANOVAs. As expected, participants in the criteria condition felt more compelled to use originality as a criterion ($M = 4.27$, $SD = 0.72$) than did participants in the no criteria condition ($M = 2.66$, $SD = 1.14$; $F(1, 49) = 40.00$, $p < .001$, partial $\eta^2 = .45$).[1] In line with these results, participants in the criteria condition found it more important to select original ideas ($M = 3.31$, $SD = 0.88$) than did participants in the no criteria condition ($M = 2.72$, $SD = 0.92$; $F(1, 49) = 5.13$; $p = .028$, partial $\eta^2 = .09$).

Also as expected, participants in the criteria condition felt more compelled to use feasibility as a criterion ($M = 4.12$, $SD = 1.11$) than participants in the no criteria condition ($M = 3.14$, $SD = 1.25$; $F(1, 49) = 9.07$; $p = .004$, partial $\eta^2 = .16$). However, participants in the criteria condition found it just as important to select feasible ideas ($M = 4.08$, $SD = 0.74$) as those in the no criteria condition ($M = 4.10$, $SD = 0.77$; $F < 1$).

Productivity

We expected no differences between conditions with regard to productivity. A 2 (selection procedure: inclusion vs. exclusion) × 2 (instructions) ANOVA revealed no significant effects (ps $> .1$; see Table 1).

[1] *This ANOVA also revealed a marginal effect of selection procedure* ($F(2, 49) = 2.93$, $p = .06$, *partial* $\eta^2 = .11$); *however, post hoc analyses showed that the differences between the inclusion, exclusion, and control conditions were nonsignificant* ($p > .1$).

Table 1. Means and standard deviations for consideration set, idea quality and satisfaction in Study 1

	Direct selection		Inclusion		Exclusion	
Measure	No criteria	Criteria	No criteria	Criteria	No criteria	Criteria
Productivity	15.22 (7.36)	14.29 (6.50)	11.90 (4.77)	18.80 (6.73)	15.00 (4.06)	14.78 (7.68)
Ideas in consideration set	–	–	4.90 (2.33)[a]	6.80 (3.16)[a]	10.90 (4.20)[b]	9.00 (7.84)[b]
Originality of generated ideas[c]	2.03 (0.46)	2.14 (0.28)	2.21 (0.47)	2.19 (0.42)	2.18 (0.45)	2.08 (0.26)
Originality of selected ideas[c]	2.11 (0.60)	2.29 (1.11)	2.10 (0.57)	2.80 (0.79)	2.20 (0.79)	2.22 (0.67)
Feasibility of generated ideas[c]	3.07 (0.48)	3.06 (0.32)	3.04 (0.29)	3.02 (0.31)	3.08 (0.42)	3.24 (0.44)
Feasibility of selected ideas[c]	3.22 (0.67)	3.14 (1.07)	3.20 (0.63)	3.40 (0.52)	3.20 (0.79)	3.00 (1.00)
Satisfaction with selection process[c]	4.44 (0.53)[d]	4.00 (0.82)[e]	4.20 (0.42)[d]	3.90 (0.99)[e]	4.60 (0.52)[d]	4.33 (0.71)[e]
Certainty of selection quality[c]	3.56 (1.24)	4.14 (0.69)	3.40 (0.97)	3.70 (1.16)	4.00 (1.16)	3.89 (0.93)

Note. $N = 55$; standard deviations are in parentheses.

[a,b]Means within a row with different superscripts are different from each other with $p < .05$.

[c]Maximum value = 5.

[d,e]Means within a row with different superscripts are different from each other with $p < .1$.

Consideration set

Because participants in the direct selection condition did not form a consideration set, we performed a 2 (selection procedure: inclusion vs. exclusion) × 2 (instructions) ANOVA with consideration set as the dependent variable, omitting the direct selection condition. As expected, participants in the exclusion condition formed a larger consideration set ($M = 10.00$, $SD = 6.09$) than participants in the inclusion condition ($M = 5.85$, $SD = 2.87$; $F(1, 35) = 7.25$, $p = .011$, partial $\eta^2 = .17$).

Idea quality

Means and standard deviations for idea quality are presented in Table 1. We performed a 3 (selection procedure) × 2 (instructions) repeated measures ANOVA with the average originality of the generated ideas and the originality of the selected idea as the within-subjects factor originality, and a similar ANOVA for the feasibility of the generated and selected ideas. These analyses revealed no significant main or interaction effects (all ps $> .1$). Thus, there were no differences between conditions with regard to the originality or feasibility of the generated ideas and the selected ideas. Furthermore, in none of the conditions was the originality or feasibility of the selected ideas significantly different from the average originality or feasibility of the generated ideas. The aggregate quality measure, finally, showed no differences between conditions with regard to the quality of the selected ideas (ps $> .1$) either.

Satisfaction

We analysed the two items that measured selection satisfaction with univariate 2 × 3 ANOVAs (see Table 1). Participants in the criteria condition were less satisfied (marginally significant) with the way in which they had made their selection ($M = 4.08$, $SD = 0.84$) than participants in the no criteria condition ($M = 4.41$, $SD = 0.50$; $F(1, 49) = 3.26$, $p = .077$, partial $\eta^2 = .06$). Participants in all conditions were equally certain that they had selected the best idea ($M = 3.75$, $SD = 1.02$; Fs < 1).

Discussion

This study replicated earlier results with regard to idea selection: participants selected ideas that were not significantly better than the average quality of their generated ideas. Both our experimental manipulations failed to improve idea selection. Although participants with exclusion instructions did form larger consideration sets, this had no effect on selection performance. More disappointingly, however, selection efficacy was not improved by specific selection instructions. Thus, selection efficacy remained low, even when participants received explicit instructions to use both originality and feasibility as selection criteria. There are three potential explanations for the failure of specific selection instructions to improve selection efficacy.

One possible explanation is that our participants found it difficult to take both originality and feasibility into account while making their selection. For example, Manske and Davis (1968) reported that participants had difficulty combining originality and practicality when generating uses for a common object. They attributed this to the strong negative correlation ($r = -.80$) between the originality and practicality of the proposed uses. However, the correlation between the average originality and feasibility of the ideas generated in Study 1 was weak ($r = -.25$, $p = .067$), so it is unlikely that

this was the main cause of our results. Nevertheless, it is possible that participants *perceived* originality and feasibility to be incompatible. Even though researchers usually define creativity as a combination of originality and feasibility, 'naive' participants may well regard originality as the single most important dimension of creative behaviour and creative products, and as incompatible with feasibility (e.g. Runco & Charles, 1993; also see Sternberg, 1985). This could also explain why participants with selection criteria were slightly less satisfied with the way in which they had made their selection: participants may have perceived the selection criteria to be paradoxical and hence impossible to satisfy. If so, selection performance might be improved by giving participants selection criteria that they do not perceive to be contradictory.

A second explanation may lie in the finding that all participants found it important to select feasible ideas. Inspection of the ideas that participants selected showed that these often concerned specific complaints or issues that were particularly important to our student participants (also see Rietzschel *et al.*, 2006). Perhaps, this is not surprising, given the brainstorming topic (possible improvements in psychology education). Thus, while originality clearly is important during a brainstorming task (and indeed is explicitly asked for in the brainstorming instructions), during idea selection participants may revert to another selection criterion that is much more important to them, namely whether they personally believe that the idea will actually constitute an improvement. From this perspective, feasibility is a necessary but insufficient condition, whereas originality is perhaps even regarded as irrelevant. If this were true, selection performance should be most strongly predicted by participants' tendency to select ideas that they feel should actually be adopted. Naturally, this criterion does not readily lend itself for a more objective assessment of selection efficacy. Instead, ratings of the likely effectiveness of an idea (i.e. the estimated likelihood that the idea will turn out to be an improvement) could provide an alternative. Thus, if this reasoning is correct, it should be expected that participants will spontaneously tend to select ideas that they feel are desirable and feasible, rather than original, and that their performance would be seen to be accurate when judged in terms of effectiveness (and probably feasibility), rather than originality.

A third possibility is that our participants used different criteria in their evaluation of originality and feasibility from those used by the external raters. In other words, they may have been quite effective in selecting the most original and feasible ideas according to their own criteria, which means that their 'poor' performance was only due to the disagreement between their criteria and those of our external raters. This would explain their apparently poor performance with no specific instructions, as well as the fact that giving specific instructions did not improve their performance. However, if this were true it would raise questions about the validity of the criteria used by participants and evaluators.

STUDY 2

In this study, participants were presented with a pre-generated set of ideas. They were then asked to give quality ratings of these ideas, and to make a selection from among them. The primary aims of this study were (1) to test the hypothesis that selection criteria can in fact improve selection efficacy if these are not perceived as contradictory, and (2) to test the hypothesis that participants' spontaneous tendency (i.e. when not provided with specific selection criteria) is to select desirable and feasible ideas, rather

than ideas that are original. Secondary aims of this study were to test once more the hypothesis that deeper processing of the available ideas would improve selection performance (using a different manipulation of processing depth), and to test for a possible lack of agreement regarding originality and feasibility.

Creativity instructions

In this study, we instructed our participants to select *creative* ideas, without mentioning originality or feasibility. If people regard originality as the most important dimension of creativity, such instructions should allow participants to focus on originality and hence enable them to improve their selection, at least with regard to that one dimension. Therefore, half of the participants in Study 2 were instructed to select the 'most creative' idea, and half of the participants were instructed to select 'the best' idea. We expected participants with creativity instructions to focus more strongly on originality while making their selection, to make a more effective (with regard to originality) selection, and to select more original ideas than participants without creativity instructions. Because the existing studies are ambiguous with regard to people's ability to effectively combine originality and feasibility considerations, we had no specific hypotheses with regard to feasibility. It was possible that the selection of original ideas would only be attained at the cost of feasibility. Alternatively, given that originality and feasibility need not be negatively correlated, participants might be able to combine the two dimensions effectively, even if they perceive them to be incompatible.

Further, in order to gain more insight into the criteria used in idea selection, we asked participants to indicate how strongly they tried to select original ideas, feasible ideas, and desirable ideas, i.e. ideas that they thought should be adopted.

Quality ratings and depth of processing

Eliciting originality and feasibility judgments, besides offering the opportunity to test for consistency in ratings, can also be an alternative way to induce deep processing of the available ideas. If the poor selection performance in Study 1 and in earlier studies was (partly) due to participants' selective attention to some ideas at the cost of other ideas, judging quality prior to making a selection should lead to a more effective selection. Thus, half of our participants were instructed to rate all ideas before making a selection, whereas the other participants immediately made a selection and rated the ideas afterwards. We expected that a beneficial effect of creativity instructions would be strongest for those participants who had rated the ideas before making a selection.

Method

Participants

Ninety undergraduate students initially participated in the experiment. Two participants encountered software problems while performing the experimental task and three participants did not follow instructions. Hence, the analyses reported below are based on 85 people (52 women, 33 men, mean age = 21.45 years). Participants received course credit or 7 euros (about 9 US dollars).

Independent variables and materials

The study had a 2 (instructions: creativity or default) × 2 (task order: select-first or rate-first) factorial design. Participants were assigned randomly to conditions, and worked individually throughout the experiment.

Participants in this study were presented with a pre-generated set of 50 ideas. These ideas had been generated in previous experiments, where participants brainstormed about possible improvements in education at the Department of Psychology. The ideas were selected to cover the entire range of originality and feasibility scores (from 'not at all' to 'very much'). The ideas were rephrased where that was deemed necessary to understand the content of the idea, and spelling and typographical errors were corrected.

To make a comparison between our own and the participants' average ratings possible, the 50 ideas were rated for originality and feasibility by a trained rater and the first author of this paper (for details about the rating scale and procedure, see Study 1). Inter-rater reliability was high: the intra-class correlations, ICC (two-way random model, consistency definition) were .82 for originality and .83 for feasibility. For each idea, we computed the average of the two-raters' scores for both originality and feasibility. Moreover, we had two-raters judge the likely effectiveness (i.e. the degree to which the idea was perceived to actually lead to improvements) of each idea. Again, inter-rater reliability was sufficient (ICC = .68); as with the other dimensions, we computed the average of the two judges' ratings for each idea.

Procedure

The experiment consisted of a rating task, a selection task, and a post-experimental questionnaire. Participants in the rate-first condition first performed the rating task and then the selection task; participants in the select-first condition first performed the selection task and then the rating task.

During the rating task, ideas were presented individually on a computer screen, and the participant was asked to give his or her rating of each idea. Ideas were to be rated on originality and feasibility. The programme was set up in such a way that one by one all ideas were rated first on one dimension, and then on the other. Both the order of the dimensions and the order of the ideas within each rating session were randomized.

For the selection task, participants received a printed list of these 50 ideas (in the selection task all participants read the ideas in the same order). In the accompanying instructions, it was explained that these ideas came from a brainstorming session about possible improvements of education at the Department of Psychology, and that their task would be to select the best four ideas from this set.[2] Half of the participants were instructed to 'select the four most creative ideas' from this set (creativity condition), and half of the participants were instructed to 'select the four best ideas' (default condition). Fifteen minutes were available for this selection task.

After completing the post-experimental questionnaire, participants were debriefed, thanked and paid.

[2] *The reason for having participants select the four best idea was that this would better approach the selection instructions as used by Rietzschel* et al. *(2006).*

Measures and dependent variables

Idea quality

We computed the average originality, feasibility, and effectiveness (as perceived by our trained raters) of the four ideas selected by each participant. Moreover, similarly to Study 1, we computed aggregate measures of idea quality: a 'traditional' aggregate measure, averaging the originality and feasibility of the selected ideas (similarly to what was done in Study 1), and a broader aggregate measure, averaging the originality, feasibility and effectiveness of the selected ideas.

Questionnaire items

Five items assessed which selection criteria participants used. These items asked participants to indicate on a five-point scale (1 = 'not at all', 5 = 'very strongly') whether they tried to select the most original idea, the most creative idea, the most feasible idea, the most desirable idea, and the best idea. One item assessed the degree to which participants were satisfied with the quality of the selected ideas.

Results

In this section, we will first address the level of agreement between our participants and our judges with regard to the originality and feasibility of the ideas. Next, we will go into the selection criteria that participants used, and their satisfaction with their selected ideas. This will be followed by analyses of selection effectiveness. Finally, we will report regression analyses testing how different aspects of selection performance are predicted by the selection criteria that participants used, when controlling for our experimental manipulations.

Agreement in ratings

To compare our participants' ratings of the 50 ideas in the idea set with those of our judges, we computed our participants' average originality and feasibility ratings of each idea contained in the set. Thus, we computed 50 average originality ratings (one for each single idea in the set) and 50 average feasibility ratings.

The intra-class correlation coefficient (average measure, two-way mixed, consistency definition) between the participants' average originality ratings and our judges' average originality ratings was .87, indicating strong agreement between our trained judges and the untrained participants. Similarly, the intra-class coefficient between the participants' feasibility ratings and our judges' feasibility ratings was .82. This makes it highly unlikely that the lack of effects of selection criteria in Study 1 was due to disagreement regarding originality and feasibility.

Selection criteria

Table 2 presents the correlations between the different selection criteria used by our participants. Three results are particularly interesting. Firstly, as expected, our participants seemed to interpret 'selecting the most creative ideas' as almost synonymous with 'selecting the most original ideas' ($r = .839$, $p < .001$). Secondly, the degree to which participants tried to select 'the best ideas' was highly correlated with the degree to which they tried to select the most feasible ideas ($r = .639, p < .001$)

Table 2. Intercorrelations between selection criteria used by participants in Study 2

	Most creative idea	Most original idea	Most feasible idea	Most desirable idea	Best idea
Most creative idea	–				
Most original idea	.839**	–			
Most feasible idea	−.447**	−.404**	–		
Most desirable idea	−.483**	−.476**	.759**	–	
Best idea	−.448**	−.403**	.639**	.696**	–

Note. $N = 85$; **$p < .001$.

and the most desirable ideas ($r = .696$, $p < .001$). Thirdly, participants' tendency to select for originality was negatively correlated with their tendency to select for feasibility ($r = -.404$, $p < .001$) and for desirability ($r = -.476$, $p < .001$). Thus, the more participants tried to select creative or original ideas, the less did they try to select feasible or desirable ideas.

The means and standard deviations of all questionnaire items are presented in Table 3. A 2 (instructions) × 2 (task order) MANOVA with the five items regarding selection criteria revealed a main effect for instructions ($F(5, 77) = 19.38$, $p < .001$, $\eta^2 = .56$) and an interaction of instructions and task order ($F(5, 77) = 2.34$, $p = .05$, $\eta^2 = .13$). Univariate analyses showed that, as expected, participants with creativity instructions reported a stronger tendency to select for originality ($M = 4.05$, $SD = 0.87$) than participants with default instructions ($M = 2.25$, $SD = 1.01$; $F(1, 81) = 78.09$, $p < .001$, partial $\eta^2 = .49$). Similarly, participants with creativity instructions reported a stronger tendency to select creative ideas ($M = 4.41$, $SD = 0.84$) than participants with default instructions ($M = 2.61$, $SD = 1.15$; $F(1, 81) = 70.22$, $p < .001$, partial $\eta^2 = .46$). This main effect was qualified by an interaction between instructions and task order ($F(1, 81) = 4.82$, $p = .031$, $\eta^2 = .056$); simple effects analysis revealed that the effect was stronger in the select-first condition ($F(1, 81) = 56.63$, $p < .001$) than in the rate-first condition ($F(1, 81) = 18.88$, $p < .001$). Participants with default instructions reported a stronger tendency to select the 'best' ideas ($M = 4.55$, $SD = 0.59$) than participants with creativity instructions ($M = 3.44$, $SD = 1.50$; $F(1, 81) = 20.46$, $p < .001$, $\eta^2 = .20$). In addition, participants with default instructions reported a

Table 3. Means and standard deviations of selection criteria used by participants and satisfaction in Study 2

	Select-first		Rate-first	
Measure	Default	Creativity	Default	Creativity
Most original idea	2.27 (1.16)[a]	3.75 (0.85)[b]	2.23 (0.87)[a]	4.33 (0.79)[b]
Most creative idea	2.73 (1.08)[a]	4.05 (0.99)[b]	2.50 (1.23)[a]	4.76 (0.44)[b]
Most feasible idea	4.09 (0.81)[a]	2.70 (1.34)[b]	3.55 (1.10)[a]	2.48 (1.60)[b]
Best idea	4.73 (0.46)[a]	3.35 (1.42)[b]	4.36 (0.66)[a]	3.52 (1.60)[b]
Most desirable idea	4.59 (0.73)[a]	3.35 (1.46)[b]	4.68 (0.57)[a]	3.10 (1.67)[b]
Satisfaction with quality of selected idea	4.32 (0.48)[a]	3.65 (1.14)[b]	4.23 (0.61)[a]	3.86 (1.06)[b]

Note. Maximum value = 5; $N = 85$; standard deviations are in parentheses.
[a,b]Means within a row with different superscripts are different from each other with $p < .05$.

stronger tendency to select desirable ideas (i.e. ideas that they would like to see implemented; $M = 4.64$, $SD = 0.65$) than participants with creativity instructions ($M = 3.22$, $SD = 1.56$; $F(1, 81) = 30.01$, $p < .001$).

In short, creativity instructions (as compared to default instructions) caused participants to give more weight to creativity and originality, and less to feasibility and desirability. This is in line with the pattern of correlations reported above.

Satisfaction

A 2 × 2 ANOVA revealed that participants with creativity instructions were less satisfied with the quality of the selected ideas ($M = 3.76$, $SD = 1.09$) than participants with default instructions ($M = 4.27$, $SD = 0.54$; $F(1, 81) = 5.71$, $p = .007$). Thus, participants who were instructed to select creative ideas were less satisfied about the quality of their selection. Again, this is in line with our earlier finding that participants tended to favour feasible ideas.

Idea quality – differences between conditions

Originality

The means and standard deviations for the quality of the selected ideas are presented in Table 4. A 2 (instructions) × 2 (task order) ANOVA with the mean originality of the selected ideas as the dependent variable showed that, as expected, participants with creativity instructions selected more original ideas ($M = 2.78$, $SD = 0.73$) than did participants with default instructions ($M = 1.96$, $SD = 0.42$; $F(1, 81) = 41.31$, $p < .001$, $\eta^2 = .34$). There were no other significant main or interaction effects (ps $> .1$). Thus, contrary to our hypothesis, the effect of creativity instructions was not stronger for participants in the rate-first condition than for participants in the select-first condition.

Because the average originality of the idea set was constant (since each participant was presented with the same set of ideas), we performed separate t tests for the participants with creativity instructions and the participants with default instructions, testing the originality of their selected ideas against the average originality of the idea set (= 2.18). These tests showed that participants with creativity instructions selected ideas that were more original than the idea set ($M = 2.78$, $SD = 0.73$; $t(40) = 5.27$, $p < .001$); in contrast, participants with default instructions selected ideas that were less original than the idea set ($M = 1.96$, $SD = 0.42$; $t(43) = -3.46$, $p = .001$). In short, these results show that creativity instructions (as compared to default

Table 4. Means and standard deviations for idea quality in Study 2

	Select-first		Rate-first	
Measure	Default	Creativity	Default	Creativity
Originality of selected ideas	1.92 (0.47)[a]	2.61 (0.69)[b]	2.00 (0.37)[a]	2.94 (0.74)[b]
Feasibility of selected ideas	2.87 (0.38)[a]	2.85 (0.34)[a]	3.13 (0.49)[b]	2.73 (0.45)[a]
Effectiveness of selected ideas	3.04 (0.55)[a]	2.58 (0.73)[b]	2.98 (0.56)[a]	2.21 (0.77)[b]

Note. Maximum value = 5; $N = 85$; standard deviations are in parentheses.
[a,b]Means within a row with different superscripts are different from each other with $p < .05$.

instructions) caused participants to make a more effective selection with regard to originality.

Feasibility

A 2 × 2 ANOVA with the mean feasibility of the selected ideas revealed a main effect of instructions ($F(1, 81) = 5.38$, $p = .023$, $\eta^2 = .06$); this main effect was qualified by an interaction between instructions and task order ($F(1, 81) = 4.18$, $p = .044$, $\eta^2 = .05$). Simple effects analysis revealed that, within the rate-first condition, participants with creativity instructions selected less feasible ideas ($M = 2.73$, $SD = 0.45$) than did participants with default instructions ($M = 3.13$, $SD = 0.49$; $F(1, 81) = 9.65$, $p = .003$); this difference was not found for the participants who first selected and then rated ideas ($F < 1$).

Two *t* tests of the feasibility of the selected ideas against the average feasibility of the idea set ($= 2.94$) showed that, within the rate-first condition, participants with creativity instructions selected ideas that were less feasible than the idea set ($M = 2.73$, $SD = 0.45$; $t(20) = -2.16$, $p = .043$), whereas participants with default instructions selected ideas that were slightly more feasible (marginally significant) than the idea set ($M = 3.13$, $SD = 0.49$), ($t(21) = 1.78$, $p = .089$). Participants within the select-first condition selected ideas that were neither more, nor less feasible than the idea set (ps $> .1$). Thus, creativity instructions led to a less effective selection with regard to feasibility than default instructions, but *only* within the group of participants who first rated the ideas for quality and then made their selection.

Effectiveness

An ANOVA revealed only a main effect of instructions ($F(1, 81) = 18.46$, $p < .001$, $\eta^2 = .19$): participants with default instructions selected ideas of higher effectiveness ($M = 3.01$, $SD = 0.55$) than participants with creativity instructions ($M = 2.39$, $SD = 0.76$). *t* tests of the effectiveness of the selected ideas against the average effectiveness of the whole idea set ($= 2.58$) showed that participants with default instructions selected ideas of higher average effectiveness than the idea set ($t(43) = 5.17$, $p < .001$); for participants with creativity instructions, the difference was not significant ($t(40) = -1.57$, $p > .1$). Thus, default instructions led to better selection with regard to idea effectiveness than creativity instructions.

Aggregate measures

An ANOVA with the traditional aggregate quality (average of the originality and feasibility of the selected ideas) revealed a main effect of task order ($F(1, 81) = 4.55$, $p = .04$, $\eta^2 = .05$), indicating that participants in the rate-first condition selected ideas of higher (aggregate) quality ($M = 2.69$, $SD = 0.31$) than participants in the select-first condition ($M = 2.56$, $SD = 0.34$). This ANOVA also revealed a main effect of instructions ($F(1, 81) = 23.36$, $p < .001$, $\eta^2 = .22$), indicating that participants with creativity instructions selected ideas of higher quality ($M = 2.78$, $SD = 0.32$) than participants with default instructions ($M = 2.48$, $SD = 0.27$). A 2 × 2 ANOVA with the broader aggregate quality (the average of the originality, feasibility, and effectiveness of the selected ideas) as dependent variable yielded no significant effects (ps $> .1$).

To summarize, creativity instructions led to the selection of ideas that were more original, but less effective (and less feasible, but only when ideas were first rated).

However, when the three different quality dimensions were combined, the different effects appeared to cancel each other out. When only originality and feasibility were combined into an aggregate measure, creativity instructions led to the selection of better ideas than default instructions.

Idea quality – the role of selection criteria

We also wanted to estimate the degree to which the quality of the selected ideas was a joint function of participants' selection criteria, when controlling for the effects of our manipulations. We therefore performed separate regression analyses for the different quality measures of the selected ideas. In these analyses, we used the selection criteria originality, feasibility, and desirability as predictors; given the pattern of correlations reported above, the tendencies to select creative ideas or the 'best' ideas were considered to be redundant. Instructions, task order, and the interaction of these two were included as control variables.

Originality

The originality of the selected ideas was positively predicted by participants' tendency to select for originality ($B = 0.22$, $SE = .054$, $t = 4.08$, $p < .001$), and negatively by their tendency to select for feasibility (although this effect was only marginally significant: $B = -0.10$, $SE = .056$, $t = -1.78$, $p = .08$) and desirability ($B = -0.16$, $SE = .059$, $t = -2.79$, $p = .007$; adj. $R^2 = .61$).

Feasibility

The feasibility of the selected ideas was negatively predicted by participants' tendency to select for originality ($B = -0.15$, $SE = .046$, $t = -3.19$, $p = .002$), and positively by their tendency to select for feasibility ($B = 0.12$, $SE = .048$, $t = 2.53$, $p = .014$; adj. $R^2 = .23$).

Effectiveness

The rated effectiveness of the selected ideas was negatively predicted by tendency to select for originality ($B = -0.23$, $SE = .070$, $t = -3.24$, $p = .002$), and positively by tendency to select for feasibility ($B = 0.16$, $SE = .073$, $t = 2.13$, $p = .036$; adj. $R^2 = .37$).

Aggregate measure (traditional)

The aggregate quality (combining originality and feasibility) of the selected ideas was only negatively predicted by tendency to select for desirability ($B = -0.11$, $SE = .035$, $t = -3.17$, $p = .002$; adj. $R^2 = .37$).

Aggregate measure (broader)

The broader aggregate quality (combining originality, feasibility, and effectiveness) of the selected ideas was negatively (but marginally) predicted by tendency to select for originality ($B = -0.05$, $SE = .026$, $t = -1.97$, $p = .052$), positively by tendency to select for feasibility ($B = 0.06$, $SE = .027$, $t = 2.17$, $p = .033$), and negatively by tendency to select for desirability ($B = -0.06$, $SE = .028$, $t = -2.20$, $p = .031$; adj. $R^2 = .06$).

In short, these results show that the tendency to select desirable ideas was never positively related to the quality of the selected ideas. In contrast, the effects of tendencies to select for originality and feasibility showed different results, depending on which quality measure was used.

Discussion

This study clarified several issues. Firstly, it showed that there was high agreement among our participants and trained judges in their ratings of originality and feasibility. The interpretation that participants perform poorly at idea selection because they see other ideas as original and/or feasible than trained raters therefore is implausible. Secondly, as in Study 1, making participants more carefully consider all their ideas prior to making a selection did not improve idea selection much, although there was a small effect indicating that participants who had rated their ideas before making a selection had a higher score on an aggregate measure combining originality and feasibility of their selected ideas.

Thirdly, the results of this study suggest that the poor selection performance found in the criteria condition of Study 1 was due to the dual nature of our selection criteria. When participants were simply instructed to select 'creative' ideas, they were more motivated to use originality (as opposed to feasibility and desirability) as a selection criterion, and indeed selected more original ideas than participants who were instructed to select 'the best' ideas. Furthermore, with regard to originality participants with creativity instructions made a more effective selection: they selected ideas that were more original than the idea set. This was not the case for participants with default instructions. Thus, giving participants creativity instructions enhanced the effectiveness of idea selection with regard to idea originality. Importantly, this did not necessarily lead them to select less feasible ideas. Instead, when both feasibility and originality were simultaneously taken into account, participants with creativity instructions performed better than those without these instructions. However, participants with creativity instructions selected ideas that were rated as less effective (i.e. less likely to have positive effects) than those without creativity instructions. Consequently, when all three dimensions were taken into account (i.e. originality, feasibility, and effectiveness), we found no differences among conditions.

Fourthly, originality clearly is not a quality dimension that people take into account spontaneously. Rather, people seem to focus on whether they believe the idea can and should be adopted. Thus, participants with default instructions (select the best ideas) selected ideas that were less original but more effective (i.e. more likely to have positive effects) than those with creativity instructions. Further, they also indicated that feasibility and desirability were more important criteria than creativity or originality. Finally, participants who were given default instructions were more satisfied with their selection than those given creativity instructions, probably because those given creativity instructions could not apply the selection criteria they would normally have preferred, and therefore felt that they made a poor selection.

In sum, poor selection performance found in previous studies may be due to the fact that participants spontaneously select ideas that they think should be adopted. Because these ideas are not necessarily original and/or feasible, when participants' performance is judged on these dimensions, they perform poorly. In the present study, originality and effectiveness of ideas were strongly and negatively correlated ($r = -.45$, $p = .001$),

which explains why instructing participants to select creative ideas led to a selection of ideas of lower effectiveness.

GENERAL DISCUSSION

While these two studies do not answer all questions with regard to idea selection, several issues seem clear. Firstly, the generation of creative ideas does not automatically lead to the selection of creative ideas (Study 1). Secondly, creative idea selection benefits most from specific selection criteria (Study 2); in contrast, enhancing the consideration that is given to the available ideas does not appear to make much difference (Studies 1 and 2). Thirdly (not surprisingly), people appear to have a strong preference for ideas they believe can and should be adopted, and (which is more interesting) seem to believe that this is incompatible with the selection of original ideas.

In short, people are able to recognize their original ideas and, given the right instructions, to select them. Furthermore, this does not necessarily mean selecting less feasible ideas: despite the perceived negative correlation between originality and feasibility, an effective combination of these dimensions is possible. However, if people do not use the 'right' selection criteria, their inclination to select feasible and desirable ideas can undo the creative potential of their initial production.

In this paper, we have described the results with regard to idea preferences in terms of a preference for feasibility and desirability, combined with the belief these are incompatible with originality. This suggests the negative conclusion that people are somehow biased against originality. However, the matter could also be approached from another direction. In a sense, originality is confounded with familiarity: original ideas are those ideas that are mentioned least frequently, and therefore presumably are least familiar. Several studies (e.g. Begg & Armour, 1991; Garcia-Marques & Mackie, 2001; Garcia-Marques, Mackie, Claypool, & Garcia-Marques, 2004) have found that people tend to judge familiar items more favourably than unfamiliar items. Thus, an idea that is not very original, but very familiar, may be judged more favourably than an original, but unfamiliar idea.

An important question is to what extent our results could be related to the nature of the brainstorming problem employed (i.e. the 'education' problem). Other idea generation research has sometimes used more esoteric problems, such as the question what the consequences would be if everybody suddenly had an extra thumb on each hand (e.g. Harari & Graham, 1975; Maginn & Harris, 1980). It is possible that such problems are more conducive to the selection of highly original ideas. For one thing, factors such as desirability and feasibility naturally are less likely to play an important role in idea selection when the problem is less realistic. However, in real life, creativity is usually required in the face of more realistic problems that people are personally involved in. In that sense, one might argue that the problem used in this paper adds to the external validity of our findings, in that the problem clearly was relevant and important to our participants. Previous research has shown that people may be less creative when generating ideas about realistic problems (Harari & Graham, 1975) or about problems that they are personally involved in (Illies & Reiter-Palmon, 2004); our results suggest that this remains a problem when ideas have to be selected.

A related question is whether originality really is as important as we make it out to be. After all, in many situations it is most important that an idea *works*. Originality may not be always necessary to solve a problem or to improve a situation. Our results

(particularly Study 2) show that people tend to take the desirability or the perceived effectiveness of ideas into account when making their selection. One might wonder whether a stronger focus on originality entails 'better' selection if this means that feasibility or effectiveness receive less attention (or priority). We believe so, for three reasons. Firstly, our results suggest that, on the whole, a stronger focus on originality does not necessarily lead to the selection of less feasible ideas. Secondly, one should keep in mind that brainstorming is specifically meant as a creativity-stimulating technique, and is required mostly when conventional problem-solving fails. In such situations, originality is not an end in itself, but a necessary means to reach the goal of finding a feasible solution outside the realm of conventional solutions. Thus, even though it is possible that the selection of a conventional idea may in practice lead to a highly effective solution, this would certainly raise the question why anyone should engage in brainstorming at all. Thirdly, although it clearly is possible to reliably rate ideas for originality, feasibility and effectiveness, it is not immediately obvious that feasibility and effectiveness ratings actually predict implementation success. If the effectiveness of creative ideas could easily and accurately be judged *a priori*, innovation probably would be an easier task than it is now. However, the problem appears to be that even experts are not able to predict which ideas or solutions will turn out to be effective or successful (Simonton, 1997).

Regardless of how much importance one attaches to the selection of original ideas, our results do illustrate that it is important to distinguish not only between performance in different stages of the creative process (idea generation and idea selection), but also between different quality dimensions (originality, feasibility, and possibly effectiveness). Clearly, participants' selection behaviour was not the same across these dimensions. In our view, this argues against the use of aggregate measures of idea quality, because these can easily cloud meaningful patterns of results.

Conclusion

To conclude, our results show that people tend to select ideas they believe in, rather than creative ideas. While this tendency may be natural enough, the whole point of brainstorming and related techniques is that there is value in creative ideas. Thus, unless attention is paid to the selection process and the implicit or explicit criteria people use, innovation is likely to suffer. Indeed, big improvements and innovative solutions will only come about when people take original ideas seriously, and are willing to take the risk to develop these further. How to achieve this seems a worthwhile avenue for further research.

Acknowledgements

This research was supported by a grant (NWO grant 410-21-010) from The Netherlands Organization for Scientific Research. We thank Benjamin de Boer for his assistance in collecting and coding the data, and three anonymous reviewers and the editor for their constructive comments. The studies reported in this paper were conducted while Dr Eric F. Rietzschel was at the Department of Work and Organizational Psychology, University of Amsterdam, and form part of his doctoral dissertation. Bernard Nijstad is now at the University of Groningen.

References

Amabile, T. M. (1996). *Creativity in context: Update to 'The Social Psychology of Creativity'*. Boulder, CO: Westview Press.

Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening of options. *Psychological Science*, *4*, 215–220.

Begg, I., & Armour, V. (1991). Repetition and the ring of truth: Biasing comments. *Canadian Journal of Behavioral Science*, *17*, 199–214.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*, 127–137.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*, 497–509.

Faure, C. (2004). Beyond brainstorming: Effects of different group procedures on selection of ideas and satisfaction with the process. *Journal of Creative Behavior*, *38*, 13–34.

Garcia-Marques, T., & Mackie, D. M. (2001). The feeling of familiarity as a regulator of persuasive processing. *Social Cognition*, *19*, 9–34.

Garcia-Marques, T., Mackie, D. M., Claypool, H. M., & Garcia-Marques, L. (2004). Positivity can cue familiarity. *Personality and Social Psychology Bulletin*, *30*, 585–593.

Harari, O., & Graham, W. K. (1975). Tasks and task consequences as factors in individual and group brainstorming. *Journal of Social Psychology*, *95*, 61–65.

Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Thousand Oaks, CA: Sage Publications Inc.

Heller, D., Levin, I. P., & Goransson, M. (2002). Selection of strategies for narrowing choice options: Antecedents and consequences. *Organizational Behavior and Human Decision Processes*, *89*, 1194–1213.

Illies, J. J., & Reiter-Palmon, R. (2004). The effects of type and level of personal involvement on information search and problem solving. *Journal of Applied Social Psychology*, *34*, 1709–1729.

Maginn, B. K., & Harris, R. J. (1980). Effects of anticipated evaluation on individual brainstorming performance. *Journal of Applied Psychology*, *65*, 219–225.

Manske, M. E., & Davis, G. A. (1968). Effects of simple instructional biases upon performance in the unusual uses test. *Journal of General Psychology*, *79*, 25–33.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30–46.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Social Psychology*, *65*, 219–225.

Nijstad, B. A., & De Dreu, C. K. W. (2002). Creativity and group innovation. *Applied Psychology: An International Review*, *51*, 400–406.

Osborn, A. F. (1953). *Applied imagination*. Oxford: Charles Scribner's Sons.

Parks, C. D., & Cowlin, R. (1995). Group discussion as affected by number of alternatives and by a time limit. *Organizational Behavior and Human Decision Processes*, *62*, 267–275.

Paulus, P. B., & Brown, V. (2003). Enhancing ideational creativity in groups: Lessons from research on brainstorming. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration* (pp. 110–136). New York: Oxford University Press.

Putman, V. L., & Paulus, P. B. (in press). Brainstorming, brainstorming rules and decision making. *Journal of Creative Behavior*.

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2006). Productivity is not enough: A comparison of interactive and nominal brainstorming groups on idea generation and selection. *Journal of Experimental Social Psychology*, *42*, 244–251.

Runco, M. A., & Charles, R. E. (1993). Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences*, *15*, 537–546.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition*, *21*, 546–556.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118.

Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, *104*, 66–89.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology*, *49*, 607–627.

Sternberg, R. J., & Lubart, T. (1999). The concept of creativity: Prospects and paradigms. In R. J. Sternberg & J. Robert (Eds.), *Handbook of creativity* (pp. 3–15). New York: Cambridge University Press.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.

West, M. A. (2002). Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups. *Applied Psychology: An International Review*, *51*, 335–387.

Yaniv, I., & Schul, Y. (1997). Elimination and inclusion procedures in judgment. *Journal of Behavioral Decision Making*, *10*, 211–220.

Received 6 February 2008; revised version received 19 December 2008