Hans Eysenck, education and the experimental approach: A meta-analysis of academic capabilities in university students

Rock Braithwaite a, Philip J. Corr b,*

a Humboldt State University, United States
b Department of Psychology, City University London, UK

ABSTRACT

Hans Eysenck had a long-established interest in the influence of individual differences on educational attainment, noting that typically personality traits and cognitive abilities are ignored in debates regarding educational policy and practice. Eysenck’s general scientific approach emphasized the importance of applying an experimental approach to answering social questions. Inspired by this perspective, in this article, we conducted a meta-analysis of the literature on (largely quasi) experimental intervention studies (N = 47, with 49 independent samples) aimed at enhancing mainly self-efficacy and self-confidence in order to influence a range of academic outcomes in university students (N = 5771). Results revealed small-to-moderate, but statistically significant, positive effects across all the outcome domains examined. There was little evidence for moderation of these effects, with quality of the study intervention the only one statistically significant (lower quality studies showing the largest effect sizes). Although our analysis shows the paucity of purely experimental studies in higher education research, the results are sufficiently clear to suggest that the study of individual differences variables are relevant in educational design and instruction. This is something Hans Eysenck told us to expect.

Hans Eysenck always had a strong interest in applying psychological principles and findings to social issues, and one of the most important of these is education (Corr, 2016). As Eysenck (1991, p. 191) stated, education is important because: “it concerns the future of our children, and hence our whole culture and civilization.” During the late 1960s, Eysenck involvement in the ‘progressive education’ debate resulted in the Black Papers that challenged the rush to ‘comprehensive education’ in the UK, which replaced the ‘Grammar School’ system that was based on intellectual merit and intelligence testing at age eleven (see Corr, 2016, pp. 181–3). Even the indomitable Margaret Thatcher, during her time as Education Secretary (1970–74), could do little to stem this political tide — indeed, under her ministership, more Grammar Schools were closed than under any other Education Secretary. When Eysenck asked her why there had been no research into the effectiveness and success of the new ‘progressive’ system, she told him that she had asked her Civil Servants to commission this but they had done nothing. It seemed then, and now, education is often a research-free zone, at least of the Eysenckian experimental type which prioritises data over dogma.

Much the same is true of higher education, where we might expect the influence of individual differences in appetite and aptitude for learning and scholastic attainment to be, at least, as marked if not more so than in early years’ education. As Eysenck (1998, p. 42; Eysenck’s italics) observed: “Children, as they grow up, increasingly choose their environment; this choice itself is driven by genetic factors. And they interpret their environment in terms of their genetic contributions. Our environment is structured by ourselves, on the basis of genetic drives.” Education at all levels would seem ripe for the type of experimental approach favoured by Eysenck to answer social questions.

In this article, we examine systematically experimental interventions at university designed to enhance self-efficacy and self-confidence, which are known to be associated with educational success. The direction of causation is not known, hence the need for the type of experimental methodology advocated by Eysenck. Specifically, we examine all extant experimental inventions, most of which are quasi-experimental, to determine: (a) the extent and quality of such research; and (b) the patterns of findings and their implications for psychological factors in university success. The focus is on those interventions designed to enhance university students’ core self-evaluations, specifically self-efficacy and self-confidence, on a range of academic outcomes (affect, cognition, knowledge, and assessment grades).

1. Education as a production function

A major challenge for university education is to identify the potential for the development of students’ capabilities, but this first requires an assessment of their initial endowments (e.g., cognitive abilities and
tolerance of frustration) which, themselves, have been shaped by previ-
ous investment (by schools and families). Only then can the factors of
the production function be sensibly specified. Via this route effective de-
sign features can be identified which allow successful investment in ed-
ucational processes to produce desired outcomes on human capital
(e.g., employability, social competence, intellectual curiosity, and more
generally good citizenship). The importance of university education is
underscored by the finding that over half of life-time human capital is
acquired after post-compulsory school investment (Heckman et al.,
1988).

Education may be profitably viewed through the lens of an
économics-based ‘production function’. A student’s individual invest-
ment has significant short/medium-term financial and opportunity
costs, but the longer-term rate of return on this investment is, on aver-
age, significant (Cunha & Heckman, 2008). Viewed from this perspec-
tive, university education may be conceived as a production function,
the efficiency of which is related to effective processes (e.g., methods
leading to the development of critical reasoning and social competence)
based on initial endowments of students (e.g., cognitive abilities, consci-
entiousness, and persistence) and preferences (e.g., time, risk and social
preferences). From this perspective, the development of university-
related capabilities represents a technology of skills formation (‘skill’
here being defined in terms of concrete and higher-level abstract com-
petencies), which represents a life-long process and applies equally to
families and firms and all other productive entities (Carneiro, Cunha,
& Heckman, 2003) including university education. Although in econom-
ic models these initial student-based endowments are often treated as
being exogenously determined (i.e., not affected by parameters of the
formal model), research now indicates that they can be fostered by
investment, especially in early childhood (Carneiro & Heckman, 2003).
Thus, student-based endowments can be developed by appropriate
investment.

Based on economic studies of early, experimentally-controlled,
interventions using samples of socially-impoverished children, targeted
programs have been shown to lead to the development of non-
cognitive, ‘soft’, skills (i.e., a portfolio of capabilities) which, then, have
been shown to impact upon future life outcomes (e.g., employment,
criminality, and social welfare) (Carneiro & Heckman, 2003). Indeed,
these soft skills explain variance in later life outcomes that is not
accounted for by cognitive ability alone (Heckman & Rubinstein,
2001), and, unlike cognitive abilities, these skills are relatively malleable
and sensitive to later life interventions (Shonkoff & Phillips, 2000).
Given that university aims to develop students’ capabilities, it would
be sensible to assume that some ‘interventions’ (e.g., methods of
teaching and personal tuition) are more effective than others, espe-
cially when it comes to non-cognitive skills (e.g., persistence, frustra-
tion tolerance, and sociability which, usually, fall under the
rubric of ‘personality’).

Closing the distance between initial endowment potential and pro-
duction of effective academic-related capabilities requires effective edu-
cational processes, which recognize both cognitive and emotional/
motivational factors. However, university educational processes are,
too often, not informed by psychological and educational knowledge;
rather, it is often assumed (but not often stated) that, by some process
of osmosis, the mere exposure to university life will inculcate appropri-
ate capabilities. This is not to say that there has not been considerable
research into effective teaching and learning processes at university;
but, it is fair to say, that much of this research has not been based on
experimentally-controlled studies which can tease apart causal and
consequential factors.

2. Psychological factors in education

The importance of using evidence-based knowledge to inform edu-
cational processes and practices at university is highlighted by (a) the
psychological challenges facing students at university, and (b) the
wide diversity of outcomes (with some students excelling, while others
fail to do so, or fail outright). This is important because the university
sector is charged by society with the task of developing the human cap-
ital of all students to the best of their abilities. But, to get at causal influ-
ences, experimental interventions are required – but these are few and
far between and most that fall under this rubric do not entail randomi-
zation of participants and are, thus, are not purely experimental in
design (of course, such research entails logistical problems, consider-
ation of ecological design, and the recognition that existing policies
may not be entirely adequate).

The main aim of this article is to summarize these interventions. Our
focus is not on instructional design, but on the wider psychological en-
vironment of students’ core self-evaluations, centred on self-efficacy
and self-confidence, which we assume permeate most learning process-
es at university (e.g., initiation of study, persistence, reaction to feed-
back, tolerance of frustration, and so on).

This section provides a summary of psychological constructs that
have been widely applied to education in general and which has
obvious relevance for higher education. It supports the view that these
psychological constructs are important in academic experience and out-
comes; and, further, this raises the possibility that they may be subject
to influence by experimental means.

2.1. Self-efficacy theory

The psychological construct of self-efficacy can be defined as a
person’s perception of his/her ability to perform successfully a behav-
ior (Sitzmann & Yeo, 2013). It is purported to influence decisions
about which behavior/s to engage in, and persistence in response to
difficulty, as well as actual task performance (Multon, Brown, & Lent,
1991). For this reason, self-efficacy theory has been highly influential
in the educational field. It is aligned with social learning and social cog-
nitive theory (Bandura & Schunk, 1981), and as such gives consideration
to the impact of social factors on self-efficacy beliefs.

Self-efficacy is composed of four factors: mastery experience, verbal
persuasion, vicarious feedback, and psychological feedback, with mas-
ter experience seen as the most important factor (Bandura & Adams,
1977). In the educational context, it lends itself well to targeted inten-
ventions (e.g., structuring effective feedback), and accounts for how
an individual’s conceptualisation of ability moderates any potential
self-efficacy effect (Wood & Bandura, 1989).

Work on self-efficacy (e.g., Bandura & Schunk, 1981) has provided
valuable insight into the successful motivation of individuals
performing tasks (particularly via an emphasis on proximal sub-
goals). Bandura (1993) provides a detailed account of the influence
which self-efficacy beliefs have on the cognitive development of an in-
dividual through cognitive, affective, motivational and selection pro-
cesses; he also gives consideration to how the psychosocial network of
influences affects performance (Bandura, Barbaranelli, Caprara, &
Pasterrelli, 1996). Self-efficacy theorists emphasize specific self-efficacy
(as opposed to an overall global measure) to predict domain specific be-
haviours (Pajares, 1996); for example, computer literacy (Compeau
& Higgins, 1995), capabilities of arthritis sufferers (Lorig, Chastain, Ung,
Shoor, & Holman, 1989), and mathematical ability (Pajares & Miller,
1995). Rodgers, Conner and Murray (2008) found behaviour-specific self-
efficacy beliefs to be superior predictors of selected academic (read-
ing 1, 30, or 100 pages) and health-related (e.g., tooth flossing everyday,
etting 5–10 servings of fruit and vegetables everyday) behaviours when
compared to other prominent types of control measures, namely per-
ceived control and perceived difficulty. Specific empirical work has fur-
ther demonstrated that self-efficacy has predictive utility for academic
outcomes across different age-groups (e.g. Bandura et al., 1996;
Caprara et al., 2008) and for work-related performance in a range of dif-
ferent study settings from simulated/laboratory based work to actual/
field work (Stajkovic & Luthans, 1998).
2.2. Self-confidence

As a concept, self-confidence is similar to, but not isomorphic with, Bandura’s concept of self-efficacy; it is recognized as crucial in psychology, education, and employability literatures, and is itself composed of specific features, (e.g., academic, inter-personal and occupational). It is commonly seen as a key determinant of how students respond to academic and employment-related opportunities and challenges. Broadly speaking, ‘confidence’ reflects those cognitive and affective processes that relate to the perceived capacity to use current capabilities to achieve some, not yet attained, desired outcome. It can be viewed as a general attitude to action-outcomes relations; and, In contrast, self-efficacy can be seen to comprise the psychological processes that enable these relations.

In contrast to individuals low in confidence, those high in this concept believe that they can reach their desired future states with the necessary personal investment (e.g., time, effort, and commitment). This self-belief has important emotional consequences which motivate necessary personal investment (e.g., time, effort, and commitment). This self-belief has important emotional consequences which motivate necessary personal investment (e.g., time, effort, and commitment). This self-belief has important emotional consequences which motivate necessary personal investment (e.g., time, effort, and commitment).

Linked to the above literature, there have been attempts to define the notion of specific academic ‘confidence’—usually defined as the belief in one’s capability to achieve some specific outcome (e.g., give a successful tutorial presentation). For example, a series of studies have developed and used the Academic Behavioural Confidence (ABC) scale (Sander & Sanders, 2003, 2006, 2009). This research applies the ABC model to self-efficacy (with the same four determining variables; see above). This scale has been refined to a four-factor (Sander & Sanders, 2009), leading to a 17-item scale that measures: confidence in grades, confidence in verbalising, confidence in study, and confidence in attendance. The authors obtained a significant correlation between ABC score and final year degree grade. This work is highly consistent with Multon et al. (1991) meta-analytical work which found a statistically significant relationship between self-efficacy and academic performance based upon a review of 39 studies.

2.4. Other constructs related to self-confidence

There are additional approaches which, whilst not directly relating to confidence, are worth noting for completeness. The learned optimism work of Seligman (1998), and the considerable body of work within social psychology on locus of control, are two key examples. There have also been notable attempts to create specific instruments for related concepts that may be useful for any study of confidence. For example, Vallarand et al. (1992) provided a measure of academic motivation, which was proposed as a key intermediate variable between specific self-efficacy and task performance. The Rosenberg Self-Esteem Scale (Rosenberg, 1965) is of potential use in accounting for the role of what is potentially a linked construct. Stark, Bentley, Lowther and Shaw (1991) provide a Student Goals Exploration Test which is also of relevance. Carroll and Garavalia (2004) and Klomegah (2007) apply a Motivated Strategies for Learning Questionnaire alongside the Self-Efficacy in Self-Directed Learning Questionnaire.

2.5. Perceived behavioural control and optimism

Various other theories supplement insights from the self-concept literature. For example, Orbell (2003) uses the concept of perceived behavioural control, which integrates material from a variety of psychological sources, particularly the theory of planned behaviour, behavioural control research and personality systems interaction theory. This literature reveals that the addition of perceived behavioural control to attitudes and norms substantially increases the explanation of variance for academic behaviours.

Perceived behavioural control shares much in common with notions of locus of control, and the two concepts have been conceptually linked by Ajzen (2002). This variable is of importance as it may be seen as the capability to monitor goal-directed behaviour under conditions of uncertainty and, also, in the absence of immediate positive reinforcement; and it entails restraint of action to control behaviour to match prevailing environment conditions.

Ruthig, Haynes, Stupinsky and Perry (2009) conducted work in a related area, establishing a link between perceived academic control and optimism (thus linking the concept to the influential model of learned optimism; Seligman, 1998). This study built on the foundations established in an earlier study by Ruthig, Perry, Hall and Hladkyj (2004), which detailed a successful intervention that targeted optimism and attributions via attribution retraining. Results showed that optimism alone was a risk factor in this particular study, the implication being that only when optimism is channelled by appropriate attributions can it manifest effectively in performance. Relevant literature relating to attributional retraining has focused on academic performance, with has produced mixed results. Examples of evidence for and against include Mitchell and Hiromi (2002) and Bridges (2001), respectively (see Gibb, Zhu, Alloy, & Abramson, 2002).

2.6. Aims of study

Given the importance of core self-evaluation, largely focused as they are on self-efficacy and self-confidence, in the development of university-related capabilities and academic outcomes, it is necessary to know whether they can, by targeted experimental intervention, be enhanced. However, the education literature is mixed. For example, Kahn and Nauta (2001) tested a social learning theory model of first-year college persistence to test precollege and first-semester college performance predictors. Contrary to their hypotheses, they did not find a significant role of first-semester self-efficacy beliefs, outcome expectations, or performance goals. In contrast, a similar finding was demonstrated within an experimental study that determined a negative relationship between self-efficacy and performance due to the likelihood of committing logic errors because of overconfidence (Vancouver, Thompson, Tischner, & Putka, 2002). Clarification of this literature seems warranted.

An alternative viewpoint is that, as core-evaluations may reflect dispositional aspects of personality and cognitive abilities, these psychological constructs are not amenable to change and are, so to speak, set in stone. This viewpoint is not consistent with the investment model of personality which emphasizes the malleability of personality traits (see Ferguson, Heckman, & Corr, 2011), and nor with the raison d’être of the role of the university in society. However, empirical evidence is needed to reveal whether, or not, this is the case.

To address the above issue, we conducted a meta-analysis of the existing literature on (largely) non-experimental interventions that focused on enhancing, in general terms, core self-evaluations, largely comprising the related constructs of self-efficacy and self-confidence. Although the number of such experimentally-controlled studies is relatively small, a sufficient number now exist to render a meta-analysis viable and conclusions potentially of relevance for instructional design and for fostering a productive psychological environment.

3. Method

The literature search and inclusion criteria are detailed in Supplementary Material, as are the coding and data extraction forms.
3.1. Outcome measures

Outcome measures for each study were coded to provide a summary treatment effect. This process involved listing corresponding measurement tools, questionnaires, subscales, and individual items that were involved in measuring student performance. Due to the variability for measuring student achievement and to provide consistency in reporting outcome measures three separate authors reviewed and grouped dependent variables according to the purpose of the measurement tool.

Categories were individually established and then discussed to select the best method to group and report student achievement. The emergent categories included: self-efficacy/confidence (specific and general measures); knowledge/learning (e.g., IQ, content knowledge questionnaires, etc.); learning strategies (e.g., metacognition, and feedback loops); anxiety; self-regulatory processes (e.g., self-monitoring, and organization, planning); motivation (e.g., goal profiles, and internal attributions); attitude/interest; self-perceptions (e.g., self-appraisal, and self-evaluation); social skills (e.g., collaboration, cohesion, and social integration); professional aspirations (e.g., career decisions); and academic attainment (e.g., tests, assignments, and GPA).

As shown in Table 1, outcome measures were grouped into three main broad domains: Process (i.e., self-efficacy, knowledge intellect, and learning strategies); Orientation (i.e., attitude/interest, self-regulation, motivation, self-perceptions, social skills, professional aspirations, and anxiety); and Performance (i.e., academic attainment) — these were reported in one instance (no pretest just posttest) and this contrasted with Process based knowledge (as represented by a ‘gain score’, or improvement e.g. reported improvements in confidence, organizational skills, etc.). Irrespective of the validity of the construction of three domains (and, of course, they are different ways to categorize the variables), statistics are reported separately for all the outcomes measures and, thus, their interpretation is not significantly affected by this classification. However, it seemed sensible to attempt some themetic organization, to reflect what might be different psychological levels. Performance is an outcome achievement measure and, thus readily stands apart from Process and Orientation; and in relation to the last two categories, Process is seen to reflect more formal mechanisms enabling the Orientation variables (e.g., Process ‘knowledge/intellect’ should be expected to influence the expression of Orientation ‘attitude/interest’, and the same may be said of ‘self-efficacy’ and ‘anxiety’, respectively).

Statistical consideration, outlier analysis, and publication bias are detailed in Supplementary Material.

3.2. Final studies

Combinations of search terms generated a possible list of 18,504 articles that were reviewed by title and abstract and reduced to a total of 371 studies. This was reduced to 47 studies with 49 independent samples meeting inclusion criteria, involving 5771 participants from 11 different countries. Results from the coding process produced an inter-rater agreement of 96.7% (r = .939) that ranged from 89.7% to 100% across the three characteristics (intervention, sample, and study features; see Supplementary Material). Based on the types of coding disagreements, there were 21 total disagreements including 9 factual errors that were corrected and 12 interpretation errors and independently reviewed and coded by a third author to determine final codes.

4. Results

We examined the effects of experimental interventions on university students’ academic-related outcomes. Descriptive information for the studies that met our inclusion criteria are given in Supplementary Material (S1 & S2). Cohen’s (1988) criteria has established that effect sizes are small (±0.20), medium (0.50), or large (≥0.80) with positive effect sizes interpreted as treatment groups having stronger results than control or comparison conditions. Negative effect sizes indicated that control groups or comparison measures yielded larger outcomes.

4.1. Random effects model

Overall, we found a small positive effect (g = 0.279, SE = 0.042, CI = 0.186, 0.361, p < .001) for university students exposed to experimental treatments. This small effect represented approximately one quarter of a standard deviation improvement on learning outcomes and achievement. The differences between the individual outcome measures (Table 1) qualify this omnibus statistic.

Analysis of homogeneity statistics determined there was a significant heterogeneous distribution (Q_{total} = 208.6, p < .001) of studies requiring subgroup analyses to explain a large portion of variance (I^2 = 78.85) between study covariates. Review of standard residuals produced six outliers (Chyung, Winiecki, & Fenner, 1998, z = 4.34; Duijnhouver, Prins, & Stokking, 2010, z = −2.50; Gaudine & Saks, 2004, z = −2.08; Latham, 2006, z = −2.48; Papinczak, Young, Groves, & Haynes, 2008, z = −2.17; Rampp & Guffey, 1999, z = 1.98), therefore, a sensitivity analysis was performed. The CMA version 2 software (Borenstein, Hedges, Higgins, & Rothstein, 2005) provides a ‘one

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Effect size statistics</th>
<th>Null test</th>
<th>Homogeneity statistics</th>
<th>Publication bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k</td>
<td>g</td>
<td>SE</td>
<td>s^2</td>
</tr>
<tr>
<td>Random effects model</td>
<td>47</td>
<td>0.279</td>
<td>0.042</td>
<td>0.002</td>
</tr>
<tr>
<td>Process</td>
<td>40</td>
<td>0.289</td>
<td>0.047</td>
<td>0.002</td>
</tr>
<tr>
<td>Knowledge/intellect</td>
<td>7</td>
<td>0.598</td>
<td>0.163</td>
<td>0.027</td>
</tr>
<tr>
<td>Learning strategies</td>
<td>15</td>
<td>0.255</td>
<td>0.076</td>
<td>0.006</td>
</tr>
<tr>
<td>Orientation</td>
<td>6</td>
<td>0.319</td>
<td>0.163</td>
<td>0.027</td>
</tr>
<tr>
<td>Attitude/interest</td>
<td>8</td>
<td>0.229</td>
<td>0.043</td>
<td>0.002</td>
</tr>
<tr>
<td>Self-regulation</td>
<td>6</td>
<td>0.087</td>
<td>0.079</td>
<td>0.006</td>
</tr>
<tr>
<td>Social skills</td>
<td>4</td>
<td>0.368</td>
<td>0.190</td>
<td>0.036</td>
</tr>
<tr>
<td>Self-perceptions</td>
<td>4</td>
<td>0.241</td>
<td>0.151</td>
<td>0.023</td>
</tr>
<tr>
<td>Professional aspirations</td>
<td>7</td>
<td>0.269</td>
<td>0.089</td>
<td>0.008</td>
</tr>
<tr>
<td>Anxiety</td>
<td>4</td>
<td>−0.134</td>
<td>0.181</td>
<td>0.033</td>
</tr>
<tr>
<td>Academic attainment</td>
<td>14</td>
<td>0.259</td>
<td>0.088</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Note. A = Total Q-value used to determine heterogeneity; B = Between Q-value used to determine significant differences between moderators; k = number of effect sizes; g = Effect size (Hedges g); SE = Standard Error. s^2 = variance. 95% CI. = Confidence Intervals (lower limit, upper limit); Z = test of the null hypothesis. χ^2 = Between study variance in random effects model. I^2 = Total variance explained by moderators. A = Total Q-value used to determine heterogeneity. Fail Safe N = number of studies needed to increase p > .05.

* p ≤ .05.
study removed program feature that completed the sensitivity analysis finding only a small change in effect size (g = 0.262) would result from removing any single study and remain close to or within the 95% confidence interval. The Fail Safe N calculation determined that an additional 1810 studies were needed to produce results that would exceed the predetermined alpha value (α = .05). Publication bias was deemed marginal, therefore, the Trim and Fill procedure was not needed to provide an unbiased estimate of overall treatment effect.

4.2. Outcome analyses

We found that experimental interventions influenced a diverse range of outcomes measures. Outcome variables not reported by more than four studies were removed from the analysis as estimates of effect size can be imprecise (Borenstein, Hedges, Higgins, & Rothstein, 2009). Overall, there were small-to-moderate treatment effects across all outcomes, ranging from −0.134 to 0.598 effect sizes.

The largest treatment effects in the Orientation domain were perceptions of self (k = 8, g = 0.368), attitude/interest (k = 6, g = 0.319), and professional aspirations (k = 7, g = 0.269). Homogeneity and publication bias statistics suggest that study distributions (Q < .05) had a high degree of variability, and that the number of studies needed to increase significant p-values beyond the threshold were suspect to publication bias.

Process variables produced the most variable treatment effects that were small-to-moderate. Objective measures of knowledge or intellect (IQ tests, content knowledge assessment, etc.) showed the largest effect sizes (k = 7, g = 0.329). Significant heterogeneity statistics (Qg = 59.13, p < .05) were indicative of a diverse distribution of study results and publication bias was improbable (Fail safe N = 124).

Performance (i.e., exam scores, assignments, or grade point averages) showed only a small overall treatment effect (k = 14, g = 0.259). Review of heterogeneity and publication bias statistics indicated consistent findings (Qg = 39.30, p > .05) for improved performance across studies with marginal publication bias (Fail Safe N = 96).

4.3. Subgroup analyses

Interpretation of the homogeneity statistics for the random effects model determined there was a heterogeneous distribution (Qtotal = 208.6, p < .001) and that a larger portion of variance (I2 = 78.85) could be explained by conducting subgroup analyses. Table 2 summarizes the moderator statistics for the coded intervention characteristics, and Table 3 the sample and study characteristics.

In summary, there were overall trends (p < .05) indicating improved learning and performance outcomes for students experiencing experimental treatments; however, study quality was the only moderating variable within intervention characteristics to produce significant differences (Qg = 6.601, p < .05) between categories. Borenstein et al. (2009) have recommended that when interpreting moderating variable differences conservative approaches should be employed when subgroups are minimal (k < 5) as estimates of treatment effect may be imprecise. We have selected to report moderator statistics and provide a conservative interpretation in order to highlight trends and recommend future directions.

4.4. Intervention characteristics

Most of the categories within intervention characteristics produced positive trends including larger effect sizes for: (a) Experimental

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention moderator statistics.</td>
</tr>
<tr>
<td><strong>Effect size descriptive statistics</strong></td>
</tr>
<tr>
<td>k</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Random effects model⁶</td>
</tr>
<tr>
<td>Intervention characteristics⁵</td>
</tr>
<tr>
<td>Design</td>
</tr>
<tr>
<td>Experimental</td>
</tr>
<tr>
<td>Quasi experimental</td>
</tr>
<tr>
<td>Duration</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Semester</td>
</tr>
<tr>
<td>Year</td>
</tr>
<tr>
<td>Follow-up</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Foundation</td>
</tr>
<tr>
<td>Atheoretical</td>
</tr>
<tr>
<td>Theoretical</td>
</tr>
<tr>
<td>Classification</td>
</tr>
<tr>
<td>Multistructural</td>
</tr>
<tr>
<td>Relational</td>
</tr>
<tr>
<td>Unistructural</td>
</tr>
<tr>
<td>Nature</td>
</tr>
<tr>
<td>Reproductive</td>
</tr>
<tr>
<td>Transformational</td>
</tr>
<tr>
<td>Intent</td>
</tr>
<tr>
<td>Primary</td>
</tr>
<tr>
<td>Secondary</td>
</tr>
<tr>
<td>Quality</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>Medium</td>
</tr>
<tr>
<td>High</td>
</tr>
</tbody>
</table>

Note. A = Total Q-value used to determine heterogeneity; B = Between Q-value used to determine significant differences between moderators; k = number of effect sizes; g = Effect size (Hedges' g); SE = Standard Error; s² = variance; 95% C.I. = Confidence Intervals (lower limit, upper limit); Z = test of the null hypothesis; I² = Between study variance in random effects model. P = Total variance explained by moderators. A = Total Q-value used to determine heterogeneity. B = Between Q-value used to determine significant (α = .01) differences between moderators.

* p ≤ .05.
produced no significant differences between subgroups. Lower quality interventions (4.5. Sample and study characteristics) only category within intervention characteristics to produce significant differences for the various experimental interventions. Overall there were consistently small-to-moderate treatment effects across all outcomes measures. There were notable effects on the process variable of knowledge/intelligence, and on the orientation variables of self-perceptions and professional aspirations. The overall effect on the performance variable of academic attainment was modest. In terms of subgroup effects, although all subgroups produced overall trends indicating improved learning and achievement outcomes for the experimental treatments, only study quality significantly moderated these effects: Lower quality interventions produced significantly greater effects for treatment groups or conditions.

4.5. Sample and study characteristics

No significant differences were found within sample variables; however, there were several trends including larger treatment effects for interventions conducted on low ability students (g = 0.339, Z = 4.113, p < .05), and treatments completed at Universities within the United States (g = 0.353, Z = 6.163, p < .05). Analyses of study characteristics produced no significant subgroup differences but, similar to previous findings, trends were present in the data. Larger treatment effects were found for unpublished studies (g = 0.305, Z = 3.064, p < .05), affective study outcomes (g = 0.359, Z = 5.937, p < .05), and measures that involved student self-reporting (g = 0.300, Z = 5.543, p < .05). In summary, sample and study subgroup variables produced small positive treatment effects.

5. Discussion

Taking our inspiration from Hans Eysenck's emphasis on the need for empirical, and preferably experimental, studies of the effectiveness of education design and instruction, we examined the effects of (largely) non-experimental interventions aimed at enhancing core self-evaluation, mainly comprising self-efficacy and self-confidence on university educational outcomes. Results revealed a wide variety of effects for the various experimental interventions. Overall there were consistently small-to-moderate treatment effects across all outcomes measures. There were notable effects on the process variable of knowledge/intelligence, and on the orientation variables of self-perceptions and professional aspirations. The overall effect on the performance variable of academic attainment was modest. In terms of subgroup effects, although all subgroups produced overall trends indicating improved learning and achievement outcomes for the experimental treatments, only study quality significantly moderated these effects: Lower quality interventions produced significantly greater effects for treatment groups or conditions.

In addition, most of the categories within intervention characteristics produced positive trends including larger effect sizes for: (a) experimental design studies employing academic year interventions; (b) multistructural interventions attempting to improve student achievement using a range of strategies or procedures; (c) interventions that used self-efficacy/self-confidence as the primary basis for improving student learning and achievement; and
(d) reproductive interventions that focused on improving student content knowledge and/or skills.

No significant differences were found within sample subgroup variables; however, there were several trends including larger treatment effects for interventions conducted on low scoring ability students and experimental treatments completed at Universities within the United States. Analyses of study characteristics produced no significant subgroup differences but larger treatment effects were found for unpublished studies, affective study outcomes, and measures that involved student self-reporting. The fact that lower quality studies, unpublished studies, and self-report yielded significant differences suggests that these variables need to be considered when interpreting the results of any one study.

5.1. Theoretical and practical implications

Although the data set was relatively small, it did yield results that hold relevant theoretical and practical implications. The first one is that it may be too easy to over-interpret the results of non-experimental studies. As shown here, even with experimentally-controlled studies, effect sizes are small and there is a negative relationship between study quality and effect size. This is a rather unfortunate outcome for evidence-based design of university education because it seems that poorly designed and conducted studies are the ones which may have the largest influence by virtue of their larger effect sizes.

Results from several outcome and subgroup analyses connect several conceptual elements of the influence that self-concepts have on university-level capabilities. Our findings may have both immediate and lasting implications as they show that increases in time (length of intervention subgroup variable) produces improved student outcomes. Furthermore, studies that performed follow-up analyses (retention measures) demonstrated that the development of capabilities remained consistent after interventions. This finding is perhaps unsurprising, but it is important: interventions need to be targeted and sustained in order for enhanced outcomes to be sustained.

One obvious implication is that programme design to foster university capabilities needs to be implemented early, and also often. Although we did not examine the possibility, it is likely that there is considerable synergy between the development of cognitive, affective and behavioural capabilities. Indeed, it is likely that this synergy takes a statistical interaction form, with each component multiplied by all others and, as such, these bundles of capabilities need to be jointly considered.

Learning is a process that has the potential to alter a student’s current level of cognition and affect, and self-efficacy and self-confidence are identified as significant predictors of several outcome variables. Results from the subgroup analyses on the intervention characteristics classification, nature, and intent provide suggestions on the implementation of targeted interventions. When considering the number of skills (intervention classification subgroup analyses) to be implemented during an intervention, our results indicate that singular or multi-component skill sets can be developed with success. Also, apparent from the results was the nature of interventions that facilitate students’ (a) specific (content) skills that produce immediate effects, or (b) more general (transformational) skills that may be expected to have continuing effects on academic outcomes. Finally, when attempting to develop students’ general core self-evaluations (intervention intent subgroup analysis) careful consideration should be given to ensuring self-efficacy is the primary focus of strategies to improve student outcomes.

One longer term outcome of such enhanced academic capabilities may be seen in employability, the success of which requires bundles of cognitive, affective and interpersonal capabilities that develop over life. Indeed, the relevance of concepts such as self-efficacy and self-confidence has been shown in a number of studies. For example, Wanberg, Zhang and Diehn (2010) position job search confidence as one of the seven key factors affecting employment outcomes; and Wanberg, Zhu and Van Hooft (2010) use the concept of re-employment efficacy for the unemployed, suggesting that this operates in a direct feedback loop with achievements relating to re-employment (see also, Knight & Yorke, 2004). The positive influence of the development of key academic capabilities should be expected to persist beyond the confines of the university campus.

5.2. Limitations

When conducting meta-analyses appropriate methods need to be employed to prevent inflated estimates of effect size. Two such concerns that have the potential to influence effect size estimates include publication bias and studies not reporting sufficient data that would permit accurate calculations of effect sizes. To address publication bias the authors established and followed a priori inclusion criteria when conducting the literature search, reported inter-rater reliability statistics for relevant information extracted from studies, and used several statistical procedures (i.e., funnel plot review, Trim and Fill procedure, and Fail Safe-N calculation) to control for publication bias. Statistical results indicated that the influence of publication bias was negligible; however, the authors recognize the possibility that studies (either published or unpublished) could have been missed during the literature search process.

Insufficient data can also influence the estimate of effect size and there were several studies that failed to report baseline information or only reported data that was significant, were ambiguous when reporting validity and reliability of the measurement tools used to collect data, and/or did not provide enough information concerning moderating variables that influence self-efficacy such as gender or ethnicity. We have attempted to control for these issues by conducting statistical analyses (i.e., subgroup and outcome analyses) and by providing interpretative precautions for an accurate perspective of the self-efficacy treatment effects on a variety of outcomes in university contexts.

Several other limitations of this literature are noteworthy. First, the majority of studies do not use pure intervention designs as they do not employ full randomization of participants – this is in the very nature of the types of studies conducted in higher education. This limits the scope of interpretation of results, which did suggest that more pure experimental designs yielded larger effect sizes. Second, studies differ in the nature of their interventions and the types of outcome measures used. This makes the task of comparison all the more difficult, requiring grouping of different measures into rather broad categories (e.g., performance attainment). Despite these problems, theoretically relevant results are still observed.

6. Conclusions

Results of our meta-analysis reveal that (albeit largely) non-experimental interventions aimed at enhancing core self-evaluations have statistically significant impacts on a range of university-related capabilities and outcomes. However, as the effect sizes were small-to-moderate, our results suggest that, as researchers and instructors, we should moderate our enthusiasm for results from studies that are not purely experimental in design and where quality of intervention is in question. This conclusion points to the need for far more pure, and higher quality, experimental studies, including ones applying interventions for longer duration, and employing follow-up measures to determine the extent of change overtime.

Our conclusions are consistent with Hans Eysenck’s emphasis on the need for rigorous empirical studies to decide the psychological dynamics of educational attainment, even at university level. As we have shown, the fragmented and inadequate nature of his literature reflects the failure to apply purely experimental approaches. As in many other
areas of social concern, Eysenck's scientific principles and practices are badly needed.

Acknowledgment

We wish to thank Dr. Victoria Scaife (School of Psychology, University of East Anglia, UK) for her contribution to an earlier draft this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.paid.2016.03.040.

References


