


# Quantifying the Effects of Motivation on Listening Effort: A Systematic Review and Meta-Analysis

Trends in Hearing  
Volume 26: 1–32  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/23312165211059982  
journals.sagepub.com/home/tia  


Peter J Carolan<sup>1,2</sup> , Antje Heinrich<sup>1,2</sup> ,  
Kevin J Munro<sup>1,2</sup>  and Rebecca E Millman<sup>1,2</sup> 

## Abstract

Motivation influences the amount of listening effort (LE) exerted or experienced under challenging conditions, such as in high-noise environments. This systematic review and meta-analysis is the first to quantify the effects of motivation on LE. The review was pre-registered in PROSPERO, and performed in accordance with PRISMA guidelines. Eligible studies examined the influence of motivation or individual traits (related to motivation) on LE in adults. Motivational factors, coded as independent variables, included financial reward, evaluative threat, perceived competence, feedback, and individual traits. LE outcomes were categorized as subjective, behavioral, or physiological. The quality of evidence was assessed using an adaptation of the Cochrane Collaboration Risk of Bias Tool. Nested random-effects meta-analyses were performed to quantify and compare the influence of motivational factors across LE outcomes. After assessing 3,532 records, 48 studies met the inclusion criteria and 43 were included in the meta-analyses. Risk of bias was high, for example, many studies lacked sample size justification. Motivational factors had a small-to-medium effect (mean Cohen's  $d = 0.34$ , range: 0.11–0.72) on LE. When LE outcomes were considered collectively, an external manipulation of motivation (perceived competence) produced a larger mean effect size compared with individual traits. Some combinations of motivational factors and LE outcomes produced more robust effects than others, for example, evaluative threat and subjective LE outcomes. Although wide prediction intervals and high risk of bias mean that significant positive effects cannot be guaranteed, these findings provide useful guidance on the selection of motivational factors and LE outcomes for future research.

## Keywords

effortful listening, financial reward, evaluative threat, individual traits, perceived competence

Received 26 April 2021; Revised received 11 October 2021; accepted 27 October 2021

## Introduction

Listening can be effortful in adverse conditions (see Mattys et al., 2012 for a review), especially if the listener has a hearing impairment (Alhanbali et al., 2017, 2018; Hornsby, 2013; Petersen et al., 2015), and this can lead to disengagement from listening and social withdrawal (Herrmann & Johnsrude, 2020). However, some individuals are motivated to exert listening effort (LE) under challenging conditions if the benefits outweigh the cost (Matthen, 2016). To date, the influence of motivation on LE has not been systematically reviewed and quantified.

The Framework for Understanding Effortful Listening (FUEL; Pichora-Fuller et al., 2016) defines LE as the *deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task*. According to FUEL, the allocation of limited-capacity

mental resources is influenced by a listener's motivation as well as other factors such as working memory capacity and fatigue. FUEL applies a general theory of motivation, Motivation Intensity Theory (Brehm & Self, 1989) to listening effort. Within MIT, effort investment is predicted to increase in line with task demands, provided there is motivation to succeed at the task and the goal is attainable.

<sup>1</sup>Manchester Centre for Audiology and Deafness, School of Health Sciences, The University of Manchester, Manchester, UK

<sup>2</sup>Manchester Academic Health Science Centre, Manchester University Hospitals NHS Foundation Trust, Manchester, UK

### Corresponding Author:

Peter J. Carolan, The University of Manchester, A4.01 Ellen Wilkinson Building, Oxford Road, Manchester M13 9PL, UK.  
Email: peter.carolan@postgrad.manchester.ac.uk



Resources are only allocated to achievable goals for which the importance of success justifies the required effort expenditure. In the context of listening tasks, MIT (Brehm & Self, 1989) predicts an interactive effect between task (listening) demands and motivation: maximal effort is invested when the listener is sufficiently motivated and listening is demanding enough to require effort but not so demanding that the effort is perceived as being wasted.

An alternative (but not mutually exclusive) model, the Cognitive Energetics Theory (CET) (Kruglanski et al., 2012), considers how an individual's ability level and personality traits influence effort. The CET is intended to apply to *all instances of goal-directed thinking*, not only *deliberate* resource allocation (cf. FUEL) and proposes that goal-directed activity results from a dynamic interaction of two forces: for goal-directed activity to occur, a "driving force" must overcome a "restraining force". The driving force consists of two main components: the individual's pool of available resources and goal importance, which is determined by the value and likelihood of attaining the goal. The magnitude of the restraining force is defined in reference to three constraining components (a) the task demands—as perceived by the individual, (b) alternative goals—which compete with the target activity for resources, and (c) resource conservation—which varies as a function of how likely the individual is to seek to withhold resources. According to CET, individual differences in personality traits affect resource conservation and hence influence effort investment. Here we review different motivational factors and different methods of measuring LE.

## Categorization of Motivational Factors

Motivation refers to the *energization and direction of behavior* (Elliot, 2006), that is, motivation underlies why people decide to behave in a certain way and how intensely they conduct that behavior. The five factors below represent different categories of motivation in experimental LE research. Four factors (financial reward, evaluative threat, perceived competence and feedback) rely on extrinsic manipulations of motivation, whereas individual traits are inherent to a person. How different motivational factors influence effect sizes when measuring the experience or expenditure of LE is currently unknown.

### Financial Reward

Consistent with predictions of cost-benefit analysis when allocating effort, previous work suggests that financial reward can motivate participants to increase LE (e.g., Carolan et al., 2021; Koelewijn et al., 2018; Mirkovic et al., 2019; Plain et al., 2020; Richter, 2016). Some studies support MIT predictions of a greater effect of motivation at higher compared to lower levels of listening demands, (e.g., Mirkovic et al., 2019; Richter, 2016), whereas other

studies did not find a significant listening demand/motivation interaction (e.g., Carolan et al., 2021; Koelewijn et al., 2021; Plain et al., 2020). Furthermore, it is difficult to determine the range of difficulty levels within which the proportional relationship between task difficulty and effort holds, according to MIT predictions. For example, in Plain et al. (2020), high levels of LE (measured using cardiac reactivity) and low self-rated giving up were measured in the most demanding listening condition, suggesting that participants continued to engage despite high levels of task difficulty, at least in the absence of performance feedback (see also Zhang et al., 2019).

### Evaluative Threat

Arousal is posited to increase the amount of resources an individual is able to allocate towards a listening task (Kahneman, 1973; Pichora-Fuller et al., 2016). Manipulations which increase arousal include evaluative threat (sometimes understood as "evaluative stress"), for example, informing participants of an upcoming performance assessment. These manipulations often include an element of social pressure, for example, in Zekveld et al. (2019) participants were informed that their data would be unusable if they failed to perform the task at a particular level. Evaluative threat may motivate greater effort investment in order to protect self-esteem from potential negative evaluations (Gendolla & Richter, 2010). Evaluative threat has been shown to increase LE (e.g., Carrillo-de-la-Pena and Cadaveira, 2000; Picou and Ricketts, 2014; Zekveld et al., 2019). Picou and Ricketts (2014) carried out a sentence recognition task using evaluative threat as a motivating factor and found differential effects depending upon the type of stimuli (audio-only vs. audio-visual) and the type of self-rated outcome measure. For audio-only stimuli, a significant increase in self-rated effort was observed in the hard SNR condition only and evaluative threat had a non-significant effect on self-rated avoidance and control (i.e., willingness to improve the situation). For audio-visual stimuli, evaluative threat increased self-rated effort and control in both SNR conditions and the likelihood of giving up decreased significantly only in the easy SNR condition. Self-rated fatigue increased significantly with evaluative threat in both the easy and hard SNR conditions for both audio-only and audio-visual stimuli.

### Perceived Competence

Competence refers to the basic need to build *skills for acting in or on the world* (Dweck, 2017). Individuals may seek to maximize their performance through greater effort investment in order to demonstrate competence (Nicholls, 1984). Increasing perceived competence experimentally may increase LE, for example, informing self-identified musicians that they should perform well on a novel auditory task because of their musical skills (McAuley et al., 2012) or

that they have been specifically selected due to their qualifications (Petrie & Carrel, 1976). Perceived competence may also be manipulated by informing participants that a listening task will be easy or hard (Hodgetts et al., 2018).

### Feedback

Participants who receive feedback are posited to increase effort to improve their performance (Kluger & DeNisi, 1996); negative feedback highlights the gap between current behavior and the desired goal (Burgers et al., 2015; Kluger & DeNisi, 1996) and positive feedback enhances feelings of competence (Cusella, 1982; Henderlong & Lepper, 2002). Generally, feedback encourages participants to set higher goals (Krenn et al., 2013; Mumm & Mutlu, 2011) and acts as a verbal reward (Deci et al., 1999), although negative feedback may prompt participants to give up if they perceive the task as too difficult (Brehm & Self, 1989). Performance feedback may influence an individual's motivation to exert LE (Gilzenrat et al., 2010).

### Individual Traits

Individual traits are potentially important factors in influencing effort and performance level as they give rise to distinctive, recurrent patterns of goal selection and pursuit (Dweck, 2017; McCabe & Fleeson, 2012, 2016). Extraversion and neuroticism may be relevant to effort regulation (Eysenck, 1990; Eysenck & Derakshan, 2011) as these traits relate to psychological needs for affiliation and social recognition, respectively (Costa & McCrae, 1988; John et al., 2008). Extraversion is associated with physiological differences that may affect effort regulation, for example lower levels of cortico-reticular arousal compared to introverts (Doucet & Stelmack, 2000; Eysenck, 1967, 1990), which may lead to greater levels of effort to compensate for suboptimal arousal conditions (Brocke et al., 1996). Neuroticism may also disrupt effort regulation (Smillie et al., 2006; van Doorn & Lang, 2010; Wallace & Newman, 1997) by directing effort to mental processes associated with negative thoughts at the expense of task-relevant activity (Wallace & Newman, 1997, 1998). Conscientiousness refers to the willingness to complete a task to the best of one's ability (Lee & Ashton, 2012) and this trait has been associated with electrophysiological measures of effortful processing, for example, the P3 component. Enge et al. (2008) found a significant effect of conscientiousness on P3 amplitude during an auditory oddball task, which may indicate increased LE (see Measures of LE below). Francis et al. (2021) included self-rated conscientiousness in a PCA dimension (along with neuroticism and openness) and found this dimension to be a significant predictor of heart rate. CET (Kruglanski et al., 2012) proposes that individual traits influence the tendency to conserve resources, for example, individuals who find mental activity enjoyable and appealing (i.e., a strong "need for

cognition") are more likely to choose resource-heavy, though potentially more successful, strategies towards goal achievement (Cacioppo & Petty, 1982; Kruglanski et al., 2012). Previous studies suggest that extraversion (Bakan, 1959; Beauducel et al., 2006; Brocke et al., 1996; Daruna et al., 1985; Ortiz & Maojo, 1993), neuroticism (Tune et al., 2018) and the need for cognition (Enge et al., 2008; Zhang, 2017) influence LE regulation.

### Measures of LE

LE may be measured using a variety of outcomes. These can be categorized as subjective (e.g., self-rated questionnaires), behavioral [e.g., performance accuracy and reaction times (RTs)] and physiological outcomes [e.g., pupillometry, cardiac reactivity, electroencephalography (EEG) and functional magnetic resonance imaging (fMRI)].

Pupil size has been shown to be sensitive to changes in arousal and cognitive load (Kahneman & Beatty, 1966; Koenig et al., 2017; Siegle et al., 2003) and thus has been proposed to be an indirect measure of LE when listening to speech under demanding listening conditions (see Winn et al. (2018) and Zekveld et al. (2018) for further details on the physiological mechanisms controlling pupil size). Another potential physiological measure of listening effort is EEG. At least one event-related potential (ERP), the P3, has been suggested to reflect effortful processing. Both the amplitude and latency of the P3 have been shown to be sensitive to the amount of attentional and memory resources required in listening tasks (Kok, 2001; Okusa et al., 1999; Polich, 2007). In addition, changes in EEG alpha band power have also been suggested to index LE (McMahon et al., 2016; Obleser & Weisz, 2012; Petersen et al., 2015; Wisniewski et al., 2017). Specifically, greater alpha activity may indicate suppression of irrelevant background noise during effortful listening tasks (Sauseng & Klimesch, 2008) or index working memory load (Petersen et al., 2015). Pupillometry and EEG are proposed to index changes in the autonomic and central nervous system related to LE, respectively (Pichora-Fuller et al., 2016).

Behavioral measures, such as RTs, may also indicate when an individual's resource capacity is under strain or exceeded, allowing inferences to be made about LE (Pichora-Fuller et al., 2016). Increased effort may manifest as slower RTs due to increased processing time (Marslen-Wilson & Tyler, 1980; Pisoni & Tash, 1974). However, how RTs respond to changes in LE may depend upon the experimental context. For instance, in a rapid judgement task, an increase in effort may result in faster RTs due to increased arousal (Carrillo-de-la-Pena & Cadaveira, 2000; Hackley & Valle-Inclán, 1998).

When multiple LE outcomes are measured simultaneously in a within-participants design, only weak correlations are found between measures, suggesting that the measures are either unreliable, or that LE is a multidimensional concept (Alhanbali et al., 2019; Herrmann & Johnsrude, 2020; McMahon et al., 2016; Miles et al., 2017; Pichora-Fuller

et al., 2016; Strand et al., 2018), or both. Moreover, even if LE outcomes are within the same category, they may not relate to the same underlying processes. For instance, results from two physiological LE outcomes (pupillometry and EEG) showed differential loading in a Factor Analysis (Alhanbali et al., 2019). FUEL recommends categorizing LE measures according to whether they are more responsive to listening demands, motivation or are responsive to both. Cardiac reactivity is proposed to be most sensitive to motivation, since it indexes the sympathetic nervous system (Brehm & Self, 1989). The FUEL suggests that behavioral measures are most responsive to listening demands, while pupillometry is hypothesized to be responsive to both listening demands and motivation (Pichora-Fuller et al., 2016).

## Aims of the Review

This review aims to evaluate all existing studies investigating motivation and LE. Within a nested analysis, we will compare effect sizes for various motivational factors on different LE outcomes. We will also compare effect sizes within categories of LE outcomes, for example, accuracy versus RTs for behavioral outcomes, and EEG versus pupillometry for physiological outcomes. It is unclear whether some combinations of LE outcomes and motivational factors produce more robust effects than others, for example, physiological outcomes may measure larger effects of individual traits on LE due to the proposed biological differences underlying traits such as extraversion (Doucet & Stelmack, 2000; Eysenck, 1967, 1990). This approach to systematically reviewing the existing literature will clarify the influence of motivational factors within a multidimensional construct of LE (Alhanbali et al., 2019; Herrmann & Johnsrude, 2020; McMahon et al., 2016; Miles et al., 2017; Pichora-Fuller et al., 2016; Strand et al., 2018).

The review aims to answer the following research questions (RQs):

- RQ1: What differences in effect sizes exist between different motivational factors?
- RQ2: What differences in effect sizes exist between different types of LE outcomes?
- RQ3: Which motivational factor/LE outcome combinations measure the largest/smallest effects?

## Methods

### Review Guidelines

The procedure and reporting of this systematic review and meta-analyses was informed by guidelines for conducting systematic reviews in healthcare, including guidance issued by the Centre for Research and Dissemination (2009), the Cochrane Handbook for Systematic Reviews (Higgins et al., 2019) and the Preferred Reporting Items for

Systematic Reviews and Meta-Analyses (PRISMA) checklist (Moher et al., 2009). The protocol was registered with PROSPERO (record number: CRD42019160953)

[https://www.crd.york.ac.uk/prospero/display\\_record.php?RecordID=160953](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=160953).

### PICOS Screening Criteria

The pre-registered PICOS (Population, Intervention, Comparator, Outcome, Study design) criteria (Table 1) were used to screen all studies. Studies were initially assessed for inclusion based on their titles. P. C. carried this out independently. R. M. independently assessed 10% of study titles and there was full agreement on suitability for inclusion. 155 studies were taken forward for screening of abstracts, which was carried out by P. C. Studies for which a decision on inclusion could not be made based on the title and abstract were screened by reading the full text. This was carried out independently by P. C. and R. M. In one case there was a discrepancy, which was resolved through discussion between P. C., R. M. and A. H.

### Assessment of Risk of Study Bias

Risk of bias was assessed using the scoring system developed by Dryden et al. (2017). This scoring system was developed based upon risk of bias assessment for clinical trials (Higgins et al., 2011) but includes only generic criteria as reporting requirements for experimental research are generally not as rigorous as those for clinical trials. Nevertheless, it is important to highlight quality issues as these reduce the reliability of effect size estimates in meta-analyses, for example, studies with low power may measure inaccurate or inconsistent effect sizes (see Caveats section). Supplementary Table S1 shows the four risk of bias questions and the score key. These questions relate to (i) sample size justification, (ii) exclusion of participant data, (iii) outcome reporting, and (iv) conflict of interest. For (i) a study was deemed high risk when no power analysis was included, unclear when complete information was not provided or low risk when a full power analysis was included. For (ii) studies at high risk excluded participants without providing details, posed an unclear risk when reporting exclusions were incomplete or low risk when details for excluding participants were fully reported. If no participants were excluded, the study was given a “not applicable” rating. For (iii) studies were deemed to be high risk when outcomes listed in the methods were not reported in the results, or unclear risk when reporting was incomplete (e.g., not including behavioral results, full statistical results or unjustified selection of a subset of parameters for a physiological outcome). Note that the research question(s) for some studies were not aligned with the aims of the present meta-analysis, which may account for bias in outcome reporting. For (iv) studies with a stated conflict of interest were high risk,

**Table 1.** PICOS Screening Inclusion and Exclusion Criteria for the Literature Search.

	Inclusion criteria	Exclusion criteria
Population	<ul style="list-style-type: none"> <li>(i) Age: 18 + or description as adults.</li> <li>(ii) Hearing: normal (either self-described or measured using pure-tone audiometry to appropriate British Society of Audiology standards) or hearing loss (with or without the use of hearing aids).</li> </ul>	<ul style="list-style-type: none"> <li>(i) Visual impairment</li> <li>(ii) Cognitive impairment</li> <li>(iii) Neurological or psychiatric disorders.</li> </ul>
Intervention	<ul style="list-style-type: none"> <li>(i) Independent variable that manipulated the participant’s motivation to invest LE and/or increase their listening task performance level. This could include tangible rewards, feedback, evaluative threat or perceived competence enhancing interventions.</li> <li>(ii) Personality trait with relevance to effort measured using a validated questionnaire (e.g., extraversion measured using the Eysenck Personality Questionnaire).</li> </ul>	Studies examining <ul style="list-style-type: none"> <li>(i) emotional valence for example stimuli that elicit strong emotions</li> <li>(ii) mood or emotional state.</li> </ul>
Comparator	<ul style="list-style-type: none"> <li>(i) Comparisons between motivated/demotivated group and a control group. Comparisons between a strong versus weak manipulation of motivation (e.g., high vs. low levels of performance-based financial reward) or between a motivated and unmotivated condition (e.g., a condition in which a performance-based reward is offered vs. passive listening).</li> <li>(ii) Comparison between groups based on personality traits (e.g., comparison between introverts and extraverts). Correlations between questionnaire scores and LE outcomes.</li> </ul>	
Outcome	Self-rated, behavioral or physiological measures of LE.	Subjective outcomes not directly related to LE (e.g., quality of life ratings). Ratings of tiredness/fatigue were excluded.
Study design	All study designs published in English in peer-reviewed or non-peer-reviewed journals, including original research data; conference papers/proceedings containing research data; book chapters; dissertations; and theses. Grey literature will also be included. The study may have been published at any date.	Exclusion criteria: Research not published in English. Case reports, systematic reviews will not be included in the analysis but may be used to source research. Interviews and clinical protocols will not be included.

unclear risk where conflicts of interest were uncertain or low risk if there were no reported conflicts of interest. All studies included in the systematic review were independently assessed for risk of bias by P. C. and R. M. In all cases, it was possible to reach agreement without the need to involve an additional author. To assess risk of publication bias, a funnel plot was generated using the dmetar package (Harrer et al., 2019) in R (R Core Team, 2018).

### Search Strategy

This review considered all relevant literature published in English to February 23, 2021. This included research

published in peer-reviewed journals, conference papers/proceedings containing research data, book chapters, dissertations and theses. A systematic search of the databases PubMed, Web of Science, Scopus, PsycINFO and Google Scholar was carried out. Search terms related to “listening effort”, “motivation” and “personality” were entered into the title, abstract and keyword fields (see Supplementary Table S2 for detailed search criteria). The search criteria were modified after registering with PROSPERO to exclude studies containing “children” in the title. After performing the search, records were transferred to Endnote and then Microsoft Excel for screening.

**Table 2.** Summary of Included Studies. This Summary Includes Sample Size, Demographic Information (Age, Hearing status), Where Available, Details of the Stimuli Used, Motivation Manipulation/Personality Trait Considered, Listening/Task Demand Manipulations, and Main Findings for Each LE Outcome. Effect Sizes (Cohen's  $d$ ), and 95% Confidence Intervals are Included for Results Included in the Meta-Analysis.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Asbjørnsen et al. (1992)</b>	36	Range: 20–30 years	NH	Syllables	None	Financial reward Reward group—high incentive \$17 and low incentive \$1.70.	1. Left ear % correct	1. For high incentive group, NS increase in left ear % correct when comparing neutral with positive $d = -0.10$ [ $-0.56$ , $0.37$ ]. For the low incentive group, NS changes in % correct when comparing neutral with positive $d = -0.19$ [ $-0.65$ , $0.26$ ].
<b>Bakan (1959)</b>	40	?	?	Digits	Single-task and dual-task conditions	Individual traits Extraversion/introversion and neuroticism (Covariate).	1. Improvement in performance (single compared to dual-task condition)	1. Sig improvement in performance for extraverts $d = 0.66$ [ $0.02$ , $1.3$ ]. NS for neuroticism (effect size not reported).
<b>Beatty et al. (1980)</b>	180	?	?	Lecture about computer music	Time compression (original 140 wpm, 210 wpm and 280 wpm)	Financial reward Course credit (amount based upon performance level) for three incentive groups, no incentive for three control groups.	1. Comprehension test—20 items, multiple choice	1. Sig greater comprehension in the incentive group. NS motivation/demand interaction.
<b>Beauducel et al. (2006)</b>	81	21.1 (2.4)	?	Tones	None	Individual traits Extraversion/introversion (EPQ), (Extreme group).	1. Performance (hits) 2. N1 amplitude 3. P3 amplitude	1. Sig greater hits for introverts $d = 0.48$ [ $0.04$ , $0.92$ ]. 2. Sig block*extraversion interaction for N1 amplitude $d = 0.4$ [ $-0.04$ , $0.84$ ]. 3. Sig block*extraversion interaction for P3 amplitude $d = 0.41$ [ $-0.03$ , $0.85$ ].
<b>Bijleveld et al. (2009)</b>	15	21	?	Digits	Recall either three or five digits	Financial reward Reward—either €0.01 or €0.50 presented visually or subliminally.	1. MPD in retention period	1. Sig greater MPD in the high compared to low reward condition $d = 0.63$ [ $0.12$ , $1.14$ ]. Reward*difficulty interaction (reward only effective in the 5-digit condition).
<b>Brocke et al. (1996)</b>	23	? Students	?	Tones	None	Individual traits Extraversion/introversion (EPI) (Extreme group).	1. P3 amplitude 2. EEG alpha power	1. Sig larger P3 amplitude for introverts $d = 0.92$ [ $0.43$ , $1.41$ ]. 2. Greater EEG alpha power for extraverts $d = 1.04$ [ $0.53$ , $1.55$ ].
<b>Cahill and Polich (1992)</b>	48	19.9 (2.2)	?	Tones	Probability of target tones (20%, 40%, 60%, or 80%).	Individual traits Extraversion/introversion (EPQ and MBTI) (Extreme group).	1. P3 amplitude 2. P3 peak latency	1. Sig personality* type*probability* gender interaction for P3 amplitude. 2. NS effect of personality on P3 latency.

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Carolan et al. (2021)</b>	24	Median 23 (range: 19–33)	NH	Sentences	Vocoded cue sentence hard (2 band) or intermediate (3 band). Vocoded target sentence always 3 band.	Financial reward (£2.50 (high) or £0.25 (low) for scoring 60% or more correct in each block.	1. Performance (accuracy) 2. RTs 3. SR work 4. SR likelihood of giving up	1. NS demand* <sup>†</sup> motivation interaction $d = 0.02$ [–0.53, 0.57] interaction for performance. 2. NS demand* <sup>†</sup> motivation interaction $d = 0.1$ [–0.44, 0.64] interaction effect for RT. 3. NS demand* <sup>†</sup> motivation interaction $d = 0.05$ [–0.48, 0.58] interaction for SR work. 4. NS demand* <sup>†</sup> motivation interaction $d = 0.17$ [–0.35, 0.69] for SR giving up.
<b>Carrillo-de-la-Pena and Cadaveira (2000)</b>	20	19 (1.37)	?	Tones	None	Evaluative threat	ERPs at Pz for target stimuli: 1a. P3 amplitude, 1b. P3 latency, 1c. N2 amplitude, 1d. N2 latency	1a. Sig larger P3 amplitudes at Pz with motivation $d = 0.63$ [0.13, 1.13]. 1b. NS effect on P3 latency $d = 0.1$ [–0.36, 0.56]. 1c. NS effect on N2 amplitude $d = 0.04$ [–0.39, 0.47]. 1d. Sig shorter N2 latencies with motivation $d = 0.6$ [0.13, 1.07]. 2. Sig faster RTs with motivation $d = 0.69$ [0.26, 1.12].
<b>Cox et al. (1999)</b>	83	Range: 60–89	HA	None	None	Individual traits MBTI	1. SR ease of communication. a. aided b. unaided c. benefit (i.e., aided—unaided)	1. NS effects of personality and Locus of Control on SR ease of communication when 1a. aided or 1b. unaided. 1c. Extroversion-introversion preference sig predictor of benefit accounting for 9.5% of variance (greater benefit with higher extroversion rating). 2. Sig faster RTs with motivation $d = 0.69$ [0.26, 1.12].
<b>Daruna et al. (1985)</b>	24	19	?	Tones	None	Individual traits Extraversion/Introversion EPI and 16PFQ (Extreme group).	1. Accuracy (rare tones) 2. Mean RT (all trials) 3. P3 amplitude (rare tones)	1. NS differences in accuracy between introverts/extraverts $d = 0.17$ [–0.63, 0.97]. 2. NS differences in RT between introverts/extraverts $d = 0.15$ [–0.65, 0.95]. 3. Sig larger P3 amplitude for introverts $d = 1.01$ [0.16, 1.86].
<b>DeWall et al. (2011) (Experiment 1A)</b>	49	?	?	Words	None	Perceived competence Three roles: manager (high power), subordinate (low power) and general worker (control).	1. Performance	1. High power group performed sig better than low power group $d = 0.93$ [0.2, 1.66] and controls $d = 0.79$ [0.07, 1.51].
<b>DeWall et al. (2011) (Experiment 2)</b>	112	?	?	Words	None	Perceived competence Power manipulation as above (Exp. 1A), after	1. Performance	1. High power group performed sig better than low power group $d = 1.06$ [0.57, 1.55] and

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>DeWall et al. (2011)</b> (Experiment 4)	165	?	?	Words	None	Perceived competence Power manipulation as above (Exp. 1A). One group were depleted, then performed either a "suitable" or "unsuitable" task for someone with power.	1. Performance	controls $d = 1.53$ [1.03, 2.03] in the self-control depletion condition. 1. Main effect of power: high power, depleted group performed worse than other groups $d = 0.38$ [0.07, 0.69].
<b>Ditraglia and Polich (1991)</b>	32	20.5 (2.2)	?	Tones	None	Individual traits Extraversion/introversion. (EPQ and MBTI) (Extreme group).	1. P3 amplitude	1. Sig personality <sup>2</sup> trial block interaction showing P3 habituated in extraverts but remained consistent in introverts $d = 1.25$ [0.49, 2.01]. 1. NS group differences in early P3a $d = -0.04$ [-0.73, 0.65]. 2. High NFCog showed larger late P3a amplitude compared to low NFCog $d = 0.75$ [0.03, 1.47].
<b>Engel et al. (2008)</b> (Study 1)	72	22.3 (2.8)	NH	Tones	None	Individual traits NFCog. (Extreme group).	1. Early P3a 2. Late P3a 3. P3b	1. NS main effect of NFCog on early P3a $d = 0.45$ [-0.16, 1.06] 2. Sig main effect of NFCog on late P3a, with larger late P3a factor scores for high NFCog $d = 0.73$ [0.11, 1.35]. sig effect of conscientiousness on late P3a $d = 0.77$ [0.15, 1.39]. 3. Sig effect of NFCog on P3b, with larger P3b amplitudes for high NFCog $d = 0.68$ [0.07, 1.3].
<b>Engel et al. (2008)</b> (Study 2)	94	22 (4.3)	NH	Tones	None	Individual traits NFCog; NEO Personality Inventory; Sensation Seeking Scale; Tridimensional Personality Questionnaire (Extreme group, controlled for adjacent components).	1. Early P3a 2. Late P3a 3. P3b	1. NS effects on NFCog on early P3a $d = 0.45$ [-0.16, 1.06] 2. Sig main effect of NFCog on late P3a, with larger late P3a factor scores for high NFCog $d = 0.73$ [0.11, 1.35]. sig effect of conscientiousness on late P3a $d = 0.77$ [0.15, 1.39]. 3. Sig effect of NFCog on P3b, with larger P3b amplitudes for high NFCog $d = 0.68$ [0.07, 1.3].
<b>Francis et al. (2021)</b>	35	37.14 (18–71)	NH	Spoken stories	Speech spoken by native English speaker masked by speech-shaped noise, non-native accented English speech	Individual traits Openness, conscientiousness, extraversion, agreeableness and neuroticism, PCA dimension 1 (extraversion, agreeableness, neuroticism), dimension 2 (openness and conscientiousness), dimension 3 (neuroticism, conscientiousness and openness).	1. NASA-TLX effort 2. Skin conductance level 3. Rate of skin conductance response 4. Heart period	1. NS effects on NASA-TLX. 2. NS effects on skin conductance. 3. PCA dimension 1 sig predictor of rate of skin conductance (main effect for both conditions) $d = 0.55$ [0.03, 1.06]. 4. PCA dimension 3 sig predictor of heart rate (main effect for both conditions) $d = 0.56$ [0.11, 1.01].

(continued)



Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Gale et al. (1972)</b>	160	Range: 18–32	?	Digits	None	Individual traits Extraversion and neuroticism (EPQ) (Covariate).	1. Mean detection scores	1. NS effects of extraversion on detection $d = -0.04$ [-0.35, 0.27] and NS correlation with neuroticism $d = 0.02$ [-0.29, 0.33].
<b>Gilzenrat et al. (2010)</b> (Experiment 2)	24	Range: 19–22	?	Tones	Low conflict condition—25% impossible discriminations High conflict condition—50% impossible discriminations	Feedback Positive or negative feedback (always received on impossible trials, depending on feedback condition).	1. RTs 2. BPD 3. PPD	1. NS trend towards longer RTs in the negative cf. positive feedback conditions $d = 0.36$ [-0.05, 0.77]. 2. NS effect on BPD $d = 0.17$ [-0.23, 0.57]. 3. NS effect on PPD $d = 0.2$ [-0.2, 0.6].
<b>Gilzenrat et al. (2010)</b> (Experiment 3)	31	Range: 19–22	?	Tones	If correct, the discrimination task was made more difficult. If incorrect, the difficulty level remained constant.	Financial reward Reward value changed depending upon difficulty level. Participants were able to reset the task. An "escape trial" was defined as the trial immediately before resetting.	Escape trials and the four trials preceding and following an escape 1. BPD 2. Expected value (calculated by multiplying the potential reward value of the trial by the expected accuracy based on previous performance) as a predictor of BPD	1. Sig inverted U-shape trend for BPD (quadratic trend analysis) $d = 1.03$ [0.59, 1.47]. 2. Expected value showed a sig quadratic trend centered on the escape trial, similar to BPD $d = 1.47$ [0.96, 1.98].
<b>Gurrera et al. (2001)</b>	18	37.8 (7.1)	NH	Tones	None	Individual traits NEO Five Factor (Covariate).	1. P3 amplitude (all sites)	1. Effect sizes when all sites are combined (significance unknown): negative correlation with neuroticism $d = 0.31$ [-0.25, 0.87], positive correlation with extraversion $d = 0.75$ [0.16, 1.34], openness $d = 0.77$ [0.17, 1.37], agreeableness $d = 0.93$ [0.33, 1.53] and conscientiousness $d = 0.65$ [0.05, 1.25].
<b>Gurrera et al. (2005)</b>	43	27.1 (9.2)	NH	Tones	None	Individual traits NEO Five-Factor Inventory (NEO-FFI) (Covariate).	1. P3 amplitude (all sites)	1. Effect sizes when all sites are combined (significance unknown): Negative correlation with neuroticism $d = 0.57$ [0.18, 0.96], positive correlation with extraversion $d = 0.47$ [0.08, 0.86], openness $d = 0.04$ [-0.38, 0.46], agreeableness $d = 0.82$ [0.44, 1.20].

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Hodgetts et al. (2018)</b>	67	27.8 (10.1)	NH	Sentences	None	Perceived competence Primed to expect an easy or hard task.	1. Mean SNR	1.2] and conscientiousness $d = 0.26$ [-0.14, 0.66]. 1. Participants in the "hard prime" group performed sig better (lower SNR) than those in the "easy prime" group $d = 0.5$ [-0.02, 1.02].
<b>Horne and Pettitt (1985)</b>	15	Range: 21–26	NH	Tones	None	Financial reward Rewards for hits, fines for misses, with increasing value over 3 days (one group sleep deprived).	1. $d'$	1. Sig greater $d'$ for sleep-deprived incentive versus no incentive $d = 1.03$ [-0.29, 2.35].
<b>Kahneman and Peavler (1969)</b>	10	?	?	Digits	None	Financial reward High financial reward (5 cents) for recall of nouns paired with odd digits, low reward (1 cent) for recall of each noun paired with even digit.	1. % correct 2. MPD: a. following digit presentation b. following response noun	1. Sig more high reward items correctly recalled than low $d = 1.5$ [0.6, 2.4]. 2a. NS effects on MPD. 2b. Sig larger MPD following high reward nouns $d = 1.00$ [0.24, 1.76].
<b>Kahneman et al. (1968) (Experiment 2)</b>	11	?	?	Digits	Short-term memory or digit transformation task (add 1 to digits heard)	Financial reward Low (2 cents) or high (10 cents) reward for correct trials.	1. Performance 2. PPD between listening and response period	1. NS effects on performance. 2. Sig task*incentive interaction— incentive had a sig effect on PPD in the short-term memory task $d = 0.91$ [0.21, 1.61] but not the digit transformation task (effect size not reported)
<b>Kang et al. (2017) (Experiment 1)</b>	19	Range: 18–26	NH	Syllables	Auditory letter was paired with either a congruent or incongruent visual letter	Financial reward (1 yuan per trial up to a maximum of 12 yuan) or no reward conditions for fast and accurate responses.	1. RTs for correct responses 2. Error rates	1. Smaller interference effect on RT in the reward condition compared to no reward for auditory targets $d = 0.56$ [0.11, 1.01]. 2. NS effect of reward on error rates for congruent targets $d = -0.05$ [-0.47, 0.37] or for incongruent targets $d = 0.13$ [-0.29, 0.55].
<b>Koelewijn et al. (2018)</b>	24	21 (median)	NH	Sentences	Presented with single-talker mask using an adaptive procedure to track 85% (easy) and 50% (hard) target intelligibility levels.	Financial reward €5 (high) or €0.20 (low) reward for 70% correct in each block of 25 trials.	1. SRT 2a. MPD b. PPD c. BPD 3a. SR effort b. SR quitting	1. NS main effect of reward on SRT $d = -0.04$ [-0.6, 0.52]. 2a. NS main effect of reward on MPD $d = 0.44$ [-0.03, 0.91]. 2b. Sig effect of reward on PPD $d = 0.46$ [-0.01, 0.92]. 2c. NS main effect of reward on BPD $d = 0.04$ [-0.4, 0.48]. 3a. NS effect of reward on SR effort $d = 0.03$ [-0.53, 0.59]. 3b. NS effect of reward on SR quitting $d = 0.2$ [-0.32, 0.72].

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Krupski et al. (1971)</b>	31	Range: 18–21	NH	Digits	Single-task or dual-task conditions	Individual traits Extraversion/introversion and neuroticism (Covariate).	1. Commissioning errors	1. Sig correlation between commissioning errors and extraversion $d = 0.7$ [0.24, 1.16]. NS for neuroticism $d = 0.24$ [–0.23, 0.71].
<b>Larisch et al. (1999)</b>	5	Range: 20–28	NH	Syllables	None	Perceived competence Participants informed that the task indexed intelligence and could differentiate successful and unsuccessful people. Highest score would receive ~\$200.	1. fMRI—contrasts assessing differential voxel activations under high vs. neutral motivation 2. Performance	1. No effect seen in 2 participants, but 3 participants showed increased BOLD response in right BA 35 and in posterior cingulate (BA 23). Decreased BOLD response seen in right BA 46. 2. NS trend towards better performance under high compared to neutral motivation $d = 0.2$ [–0.69, 1.09].
<b>McAuley et al. (2012)</b> (Experiment 1)	36	?	NH	Tones	None	Perceived competence Self-identified musicians were primed with a positive stereotype. Gains condition (awarded points for correct answers), losses condition (deducted points for incorrect answers).	1. Improvement in performance across 8 trials	1. Sig improvement for musicians compared to non-musicians in the gains condition $d = 0.65$ [0.07, 1.23]. In the loss condition, improvement for non-musicians was NS better than for musicians $d = 0.07$ [–0.5, 0.64].
<b>McAuley et al. (2012)</b> (Experiment 2)	58	?	NH	Tones	None	Perceived competence Participants primed for promotion or prevention focus.	1. % correct	1. Promotion-focus participants performed sig better in the gain condition compared to the losses condition $d = 1.16$ [0.6, 1.72]. Prevention-focus participants did not perform sig better in the losses condition compared to the gain condition $d = 0.22$ [–0.3, 0.74].
<b>Mirkovic et al. (2019)</b>	31 (15—HI; 16 NH)	HI—74 NH—66.9	HA/NH	Audiobook	Omnidirectional (low SNR) or directional (high SNR) microphone hearing aid simulator	Financial reward €20 reward offered in one of the two sessions.	1. % correct on comprehension task 2. EEG impulse responses	1. Sig effect of reward on % correct at low SNR (both groups) $d = 0.76$ [0.45, 1.07]. Sig main effect on % correct for HI $d = 0.78$ [0.37, 1.19]. NS main effect on % correct for NH $d = 0.44$ [–0.13, 1.01]. 2. NS effect of reward on EEG impulse response.
<b>Ortiz and Majojo (1993)</b>	20	19 (both groups)	?	Tones	None	Individual traits Extraversion (EPI) (Extreme group).	1. P3 amplitude 2. P3 latency	1. Sig greater P3 amplitude in introverts $d = 2.61$ [1.74, 3.48]. 2. NS difference in P3 latency between groups $d = 0.47$ [–0.17, 1.11].

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Petrie and Carrel (1976)</b>	100	?	?	20-min lecture on TV and radio speaking	None	Perceived competence Participants informed they were chosen for special training based on their qualifications.	1. Performance on multiple choice comprehension test	1. Sig higher comprehension scores in the motivated group $d = 0.51$ [0.11, 0.91].
<b>Picou and Ricketts (2014) (Experiment 1)</b>	16	24.87	NH	Sentences (audio only)	Background cafeteria noise presented at easy and hard SNRs (corresponding to approx. 50% and 80% correct in pilot testing)	Evaluative threat Participants informed of an upcoming comprehension test (motivated) or were not tested (unmotivated).	1a. SR effort 1b. SR avoidance 1c. SR control	1a. Sig effect on SR effort for hard SNR $d = 0.6$ [0.11, 1.09]. 1b. NS effect on SR avoidance for hard SNR $d = 0.28$ [-0.34, 0.9]. 1c. NS effect on SR control for hard SNR $d = 0.26$ [-0.36, 0.88].
<b>Picou and Ricketts (2014) (Experiment 2)</b>	16	23.44	NH	Sentences (audio plus visuals)	Background cafeteria noise presented at two SNRs (hard SNR 50% and easy SNR 80%)	Evaluative threat Participants informed of an upcoming comprehension test (motivated) or were not tested (unmotivated).	1a. SR effort 1b. SR avoidance 1c. SR control	1a. Sig effect on SR effort for hard SNR $d = 0.68$ [0.19, 1.17]. 1b. NS effect on SR avoidance for hard SNR $d = 0.42$ [-0.14, 0.98]. 1c. Sig effect on SR control for hard SNR $d = 0.53$ [-0.01, 1.07].
<b>Plain et al. (2020)</b>	32	22.22 (3.03)	NH	Sentences	Six SNRs	Financial reward High (€5) or low (€0.20) reward for scoring >70% correct in each block.	1. Performance 2. PEP 3. SR effort 4. SR quitting	1. NS reward*SNR interaction for performance $d = 0.16$ [-0.32, 0.64]. 2. NS reward*SNR interaction for PEP $d = 0.15$ [-0.33, 0.63]. 3. NS reward*SNR interaction for SR effort $d = 0.07$ [-0.41, 0.55]. 4. NS reward*SNR interaction for SR quitting $d = 0.14$ [-0.34, 0.62].
<b>Polich and Martin (1992)</b>	54	19.4 (1.1)	?	Tones	Target stimulus probability either .20, .50 or .80	Individual traits Extraversion/introversion, sensing/seeing, thinking/feeling, judgment/perception (MBTI), extraversion, neuroticism, psychoticism, Lie (EPQ) (Covariate).	1. P3 amplitude 2. P3 latency	1. For the whole group, NS correlations between and P3 amplitude and personality variables (extraversion/introversion $d = -0.02$ [-0.4, 0.36]; sensing/intuiting $d = 0.24$ [-0.12, 0.6]; thinking/feeling $d = 0.54$ [0.19, 0.89]; judgement/perception $d = 0.18$ [-0.18, 0.54]; extraversion $d = 0.06$ [-0.31, 0.43]; neuroticism $d = 0.47$ [0.12, 0.82]; psychoticism $d = 0.45$ [0.1, 0.8]; Lie $d = -0.35$ [-0.77, 0.07]). 2. For the whole group, NS correlations between P3

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Pritchard (1989)</b>	37	Male 23.8 (Range: 19.17–30.25) Female 24.4 (Range: 18.75–30.92)	?	Tones	None	Individual traits EPQ (all scales) (Covariate).	1. P3 amplitude 2. P3 latency	latency and personality variables (extraversion/introversion $d = -0.26$ [-0.62, 0.1]; sensing/intuition $d = -0.02$ [-0.4, 0.36]; thinking/feeling $d = -0.14$ [-0.53, 0.25]; judgement/perception $d = 0.26$ [-0.1, 0.62]; extraversion $d = 0.02$ [-0.36, 0.4]; neuroticism $d = -0.45$ [-0.89, -0.01]; psychoticism $d = 0.04$ [-0.33, 0.41]; Lie $d = 0.49$ [0.14, 0.84]). 1. For males, sig correlation between Lie scale and P3 amplitude. 2. For males, sig inverse correlation between P3 latency and neuroticism.
<b>Rammesayer and Stahl (2004)</b>	28	25.7 ( $\pm$ 5.7 years)	?	Tones	None	Individual traits EPQ (extraversion) (Extreme group).	1. Response errors 2. RT 3. NI latencies 4. NI amplitude 5. P3 latency 6. P3 amplitude 7. S-LRP latency	1. NS differences in response errors between introverts and extraverts $d = 0.2$ [-0.54, 0.94]. 2. NS differences in RT $d = 0.17$ [-0.57, 0.91]. 3. NS differences in NI latency $d = 0.08$ [-0.66, 0.82]. 4. Sig larger NI amplitude for introverts $d = 0.92$ [0.14, 1.7]. 5. Sig slower P3 latency for extraverts $d = 0.88$ [0.1, 1.66]. 6. NS group differences in P3 amplitude $d = 0.47$ [-0.28, 1.22]. 7. NS group differences om S-LRP latency $d = 0.24$ [-0.5, 0.98].
<b>Richter (2016)</b>	16	23.88	?	Tones	Sine waves presented in pairs differing by 20 Hz (easy condition) or 3 Hz (hard condition)	Financial reward \$0.20 (low) or \$2 (high) reward for scoring >90% correct in each block of 25 trials.	1. Cardiac reactivity: 1a. PEP 1b. HR 1c. SBP 1d. DBP 2. RT 3. % correct	1a. Sig effect of reward on PEP for hard condition $d = 2.97$ [2.12, 3.82]. 1b. Sig effect of reward on HR for hard condition $d = 1.00$ [0.37, 1.63]. 1c. NS effect of reward on SBP for hard condition $d = 0.39$ [-0.24, 1.02]. 1d. NS effect of reward on DBP for hard condition $d = 0.18$ [-0.43, 0.79]. 2. NS effect of reward on RT $d = 0.17$ [-0.45, 0.79].

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Schmidt et al. (2004)</b>	74 (36 Introverts and 38 Extraverts)	21.7 (2.4)	?	Tones	None	Individual traits extraversion/introversion (Extreme group).	1. RT to targets 2. Hits	3. NS effect of reward on % correct $d = 0.30$ [−0.20, 0.80]. 1. NS main effect on RT $d = 0.08$ [−0.38, 0.54]. Sig block <sup>2</sup> personality interaction for RT $d = 0.43$ [−0.03, 0.89]. 2. NS difference in hits $d = 0.09$ [−0.37, 0.55].
<b>Sewell (1985)</b>	120	19.7	?	Lecture (subject unknown)	None	Financial reward Orientating instructions: either "interesting", "boring" or no orientation; Financial reward (up to \$2.50 based on performance).	1. 20-item comprehension test	1. Orientating instructions and financial incentive had NS effect on comprehension score.
<b>Snyder and Snyder (1956)</b>	41	19.8	?	Audiobook	None	Financial reward Participants rewarded for recall from the target speaker only (but were unaware of this).	1. Recall of words	Sign test showed sig more recall for sentences spoken by the target speaker $d = 0.49$ [0.17, 0.81].
<b>Stanners et al. (1979) (Experiment 2)</b>	36 Undergraduates	?	?	Digits	Digit transformation—add either 0, 1 or 3 to each digit	Financial reward (5 cents per trial).	1. MPD 2. HR 3. Skin conductance 4. EMG	1. NS effect of reward on MPD (effect size not reported). 2. Sig effect of reward on HR $d = 0.43$ [0.09, 0.77]. 3. NS effect of reward on skin conductance (effect size not reported). 4. NS effect of reward on EMG (effect size not reported).
<b>Strand et al. (2018)</b>	111	Range: 18–28	NH	Words, digits, sentences	Speech presented at 65 dB, speech-shaped noise presented at 60 (easy)/67 dB (hard) depending upon condition)	Individual traits Extraversion—Big Five Inventory; sensitivity—HSPS (Covariate).	1. SR effort (NASA-TLX) 2. RTs (complex dual task and semantic dual task) 3. Recall: CSCT (updating/inhibiting) in high load and low load conditions 4. PPD	1. NS associations between SR effort and extraversion $d = -0.28$ [−0.57, 0.01]. HSPS $d = 0.56$ [0.32, 0.8]. 2. NS associations between extraversion and complex dual task RT $d = -0.1$ [−0.37, 0.17] and semantic dual task RT $d = 0.1$ [−0.16, 0.36]. HSPS and complex dual task RT $d = 0.75$ [0.51, 0.99] and semantic dual task RT $d = 0.63$ [0.39, 0.87]. 3. NS associations between extraversion and CSCT (updating) $d = 0.04$ [−0.22, 0.3]. CSCT (inhibiting) $d = 0.02$ [−0.24, 0.28]; CSCT (low load) $d = 0.32$ [0.07, 0.57]; CSCT (high load) $d = -0.1$ [−0.37, 0.17]. NS associations between

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Tune (1966)</b>	40	Old (Introverts 65.9, Extraverts 65.3) and young (Introverts 35.5, Extraverts 38.1)	?	Digits	None	Individual traits Extraversion/introversion (Heron inventory) (Extreme group—median split)	1. Correlation between Heron score and $d'$ on the task	HSPS and CSCT (updating) $d = -0.24$ [-0.52, 0.04]. CSCT (inhibiting) $d = -0.16$ [-0.44, 0.12]. CSCT (low load) $d = -0.3$ [-0.59, -0.01]. CSCT (high load) $d = -0.14$ [-0.41, 0.13]. 4. NS associations between PPD and extraversion $d = -0.2$ [-0.48, 0.08] and HSPS $d = 0.08$ [-0.18, 0.34]. 1. NS positive Spearman's correlation between extraversion score and $d'$ ; $d = 0.53$ [0.13, 0.93].
<b>Tune et al. (2018)</b>	29	56 (median)	NH/HI	Digits	None	Individual traits Agreeableness, conscientiousness, extraversion, neuroticism, openness (Big five) (Covariate)	Correlations between personality subdomains and parameters: 1. Performance 2. EEG 2a. AMI alpha 2b. Amplitude ALI 0.8 Hz 2c. Amplitude ITPC 0.8 Hz 2d. ITPC	1. Performance Agreeableness $d = 0.28$ [-0.54, 1.12] (NS), openness $d = -0.14$ [-0.98, 0.68] (NS), extraversion $d = 0.37$ [-0.43, 1.25] (NS), neuroticism $d = 0.65$ [-0.16, 1.58] (NS), conscientiousness $d = 0.28$ [-0.52, 1.54] (NS). 2a. AMI alpha: Agreeableness $d = 0.56$ [-0.26, 1.46] (NS), openness $d = 0.22$ [-0.58, 1.07] (NS), extraversion $d = 0.20$ [-0.61, 1.04] (NS), neuroticism $d = 0.06$ [-0.77, 0.90] (NS), conscientiousness $d = 0.08$ [-0.72, 0.93] (NS). 2b. Amplitude ALI 0.8 Hz: Agreeableness $d = 0.10$ [-0.72, 0.93] (NS), openness $d = 0.04$ [-0.77, 0.87] (NS), extraversion $d = 0.24$ [-0.58, 1.09] (NS), neuroticism $d = 1.22$ [-0.17, 2.27] (sig), conscientiousness $d = 0.28$ [-0.54, 1.12] (NS). 2c. Amplitude ITPC 0.8 Hz: Agreeableness $d = 0.36$ [-0.45, 1.22] (NS), openness $d = 0.85$ [0.02, 1.81] (NS), extraversion $d = -0.37$ [-1.22, 0.45] (NS), neuroticism $d = -0.43$ [-1.32,

(continued)

Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Weis et al. (2013)</b>	26	24 (2.0)	NH	Tones	None	Financial reward Reward session—gain low (€0.10) or high (€0.50) reward for correct answers. Punishment session—start with €25 and lose either a low (€0.10) or high (€0.50) amount for incorrect answers.	1. % correct 2. RTs 3. fMRI during reinforcement period	0.37] (NS), conscientiousness $d = -0.02$ [-0.85, 0.80] (NS). 2d. Phase delay /TPC: Agreeableness $d = 0.16$ [-0.65, 1.01] (NS), openness $d = 0.18$ [-0.63, 1.01] (NS), extraversion $d = 0.18$ [-0.63, 1.01] (NS), neuroticism $d = -0.65$ [-1.58, 0.16] (Sig), conscientiousness $d = 0.08$ [-0.75, 0.93] (NS). 1. NS effect of type of session (reward/punishment), NS effect of value, NS session/value interaction for % correct. 2. Sig session/reward interaction for RTs: punishment session $d = 0.24$ [-0.21, 0.69]; reward session $d = -0.08$ [-0.53, 0.37] 3. NS effects of reward value for BOLD response in auditory cortex. Higher BOLD response for high versus low reinforcement in the right and left visual cortex, anterior cingulate cortex and right insular.
<b>Zekveld et al. (2019)</b>	63 (34 NH; 29 HI)	52	NH and HI	Sentences	Interfering speech masker—male speaker: SRT50% and SRT71% conditions.	Evaluative threat (impossible to reach target).	1. SRT50 2. BPD 3. PPD 4. MPD 5a. SR effort 5b. SR giving up	1. NS effect of evaluative threat for SRT50: NH $d = -0.16$ , [-0.49, 0.18]; HI $d = 0.16$ [-0.21, 0.52]. 2. NS main or interaction effects for BPD: NH at SRT50 $d = -0.02$ [-0.70, 0.65]; HI at SRT50 $d = 0.32$ [-0.41, 1.06] 3. Sig main effect of evaluative threat group on PPD: NH at SRT50 $d = 0.10$ [-0.57, 0.77]; HI at SRT50 $d = 0.12$ [-0.61, 0.85]. 4. NS main effect of evaluative threat group or interactions for MPD: NH at SRT50 $d = 0.10$ [-0.57, 0.77]; HI at SRT50 $d = 0.15$ [-0.58, 0.88] 5a. Sig main effect of evaluative threat on SR effort NH $d = 1.16$ [0.43, 1.89]; HI $d =$

(continued)



Table 2. Continued.

Author(s)	Sample Size	Age Mean (SD) unless stated	Hearing level	Stimuli	Listening / task demand manipulation	Motivation/personality variable	Outcome measure(s)	Finding(s)
<b>Zhang (2017)</b>	35	22.43 (2.39)	NH	Sentences	Speech rate (slow, normal, slightly fast, fast, extremely fast)	Financial reward Reward point value of each sentence (1, 3, 5, 7 or 9) (1,000 points needed to receive reward). NFCog (Covariate).	1a. Correlation between effort efficiency and NFCog score 1b. Correlation between performance accuracy and NFCog score	0.84 [0.07, 1.60]. 5b. Sig main effect of evaluative threat on SR giving up: NH $d = 0.82$ [0.11, 1.51], HI $d = 2.50$ [1.50, 3.47]. 1a. Sig negative correlation between effort efficiency and NFCog $d = -0.88$ [-1.52, -0.24]. 1b. NS correlation between performance accuracy and NFCog $d = -0.08$ [-0.55, 0.39].
<b>Zhang et al. (2019)</b>	35	22.43	NH	Sentences	Speech rate (slow, normal, slightly fast, fast, extremely fast)	Financial reward Reward point value of each sentence (1, 3, 5, 7 or 9) (1,000 points needed to receive reward).	1. Accuracy 2. PPD: 2a. cue window 2b. sentence presentation 2c. response window	1. Sig reward/speech rate interaction for task accuracy $d = 0.26$ [-0.06, 0.58]. 2a. Sig effect of reward on PPD in cue window $d = 0.53$ [0.2, 0.86]. 2b. Sig effect of reward on PPD during sentence presentation $d = 0.52$ [0.19, 0.85]. NS reward*demand interaction for PPD $d = 0.08$ [-0.23, 0.39]. 2c. Sig demand*reward interaction for PPD during response window $d = 0.48$ [0.15, 0.81].

Notes:— not stated, 16PFQ = Sixteen Personality Factor Questionnaire, ALI = alpha lateralization index, AMI = attentional modulation index, BA = Brodmann area, BOLD response = blood oxygenation level dependent response, BPD = baseline pupil diameter, CSCT = Cognitive Spare Capacity Test, DBP = diastolic blood pressure, EEG = electroencephalogram, EMG = electromyography, EPQ = Eysenck Personality Questionnaire, ERPs = event related potentials, fMRI = functional magnetic resonance imaging, HA = participants with hearing aids, HI = participants with hearing impairment, HR = heart rate, ITPC = intertribal phase coherence, MBTI = Myers-Briggs Type Indicator, MPD = mean pupil diameter, NI = negative ERP component (peak between 80 and 120 ms after stimulus onset), NZ = negative ERP component (peak between 200 and 300 ms after stimulus onset), NASA-TLX = NASA Task Load Index, NS = non-significant, NFCog = Need for Cognition, NH = normal hearing participants, P3 = positive ERP component (peak between 250 and 500 ms after stimulus onset), PEP = pre-ejection period, PPD = peak pupil diameter, Pz = midline parietal electrode, RT = reaction time, SBP = systolic blood pressure, Sig = significant, SNR = signal-to-noise ratio, SR = self-rated, SRT = Speech reception threshold, wpm = words per minute.

## Meta-Analysis Strategy

Cohen's  $d$  was used as an effect size metric, which represents the standardized mean difference score in LE between high and low motivation conditions (either repeated-measures or between-groups comparisons). A positive Cohen's  $d$  represents an increase in LE in response to motivation. Effect sizes were calculated directly using means, standard deviations and the sample size, where this information was provided by the author(s). Where this was not possible  $F$  ratios,  $t$  tests or correlation coefficients were used. For seven studies (Asbjörnsen et al., 1992; Bijleveld et al., 2009; DeWall et al., 2011; Ditraglia & Polich, 1991; Hodgetts et al., 2018; Mirkovic et al., 2019; Picou & Ricketts, 2014), means and standard deviations were estimated from figures using WebPlotDigitizer (Rohatgi, 2020). Where insufficient information was included in published articles, attempts were made to contact the author(s) for further information. Where studies had "easy" and "hard" listening demands (or other task difficulty manipulations), the effect size was based on the interactive effect between listening demands and motivation. If this information was not provided by the author(s), the effect size for the most demanding listening condition was used. This approach was necessary as (i) there may be null results for low-demand conditions due to low LE requirements and (ii) the effect of motivation on LE is posited to be greater for higher listening demands (Brehm & Self, 1989). Signs of effect sizes were reversed for the individual trait neuroticism as higher neuroticism was expected to lower LE (see Categorization of motivational factors: Individual traits). As RTs may feasibly accelerate or decelerate with greater effort (see Measures of LE), the absolute value of the effect size was taken for RTs.

The Measure of the Effect (MOTE) package (Buchanan et al., 2017) in R (R Core Team, 2018) was used to calculate effect sizes and 95% confidence intervals (CI) for each study. For repeated-measures studies, the following equation (adapted from Lakens, 2013) was used to calculate Cohen's  $d$ :

$$d_{m} = \frac{M_1 - M_2}{\sqrt{(SD_1^2 + SD_2^2) - (2 \times r \times SD_1 \times SD_2)}} \times \sqrt{2 \times (1 - r)} \quad (1)$$

where  $M_1$  and  $M_2$  refer to the means for each motivation condition,  $SD_1$  and  $SD_2$  are the standard deviation of the mean for the unmotivated and motivated conditions, respectively, and  $r$  is the Pearson's correlation coefficient between conditions. Where the correlation coefficient was not provided (as is often the case in experimental designs), the methods outlined by Hullett and Levine (2003) were used to estimate the correlation coefficient from the  $t$ -statistic ( $t$ ) and  $F$ -statistic ( $F$ ) (where  $df_{\text{error}}$  is the degrees of freedom for the error term):

$$r = \sqrt{\frac{t^2}{t^2 + df}} \quad (2)$$

$$r = \sqrt{\frac{F}{F + df_{\text{error}}}} \quad (3)$$

If  $F$  and  $t$  were not provided, an alternate equation (Cumming & Calin-Jageman, 2017), which does not require the correlation coefficient, was used to calculate Cohen's  $d$  (where  $M_{\text{diff}}$  is the difference in the means between motivation conditions):

$$d_{av} = \frac{M_{\text{diff}}}{\left(\frac{SD_1 + SD_2}{2}\right)} \quad (4)$$

The following formula (Rosenthal, 1991) was used to calculate Cohen's  $d$  from  $t$  where means were unavailable for repeated-measures studies (where  $n$  is the sample size):

$$d_z = \frac{t}{\sqrt{n}} \quad (5)$$

For between-groups studies, Cohen's  $d$  was calculated as follows (Cumming, 2012) (where  $SD_{\text{pooled}}$  is the pooled standard deviation):

$$d_s = \frac{M_{\text{diff}}}{SD_{\text{pooled}}} \quad (6)$$

Where means were not available, the formula adapted from Lakens (2013) was used (where  $N$  is the sample size):

$$d_s = \frac{t}{\sqrt{N}} \quad (7)$$

For studies involving correlations, the 95% CI for Pearson's correlation coefficients were calculated using the *Meta-Essentials* workbook (Suurmond et al., 2017). To aid comparison with other studies, these values were converted to Cohen's  $d$  using a web-based effect size calculator (Lenhard & Lenhard, 2016).

A random-effects meta-analysis was carried out using the Meta package (Schwarzer, 2007) in R (R Core Team, 2018). Due to potential covariates not considered here (e.g., participants hearing level), a random-effects rather than a fixed-effects meta-analysis model was applied (Thompson & Higgins, 2002). A summary effect size (Cohen's  $d$ ), 95% CI around this effect size,  $Z$ -statistic and  $p$ -value were calculated to provide a test of whether the summary effect size differed significantly from zero. The inconsistency measure,  $I^2$ , and its 95% CI was calculated as a measure of heterogeneity, expressed as a percentage (Higgins et al., 2019). Prediction intervals (PIs) were calculated using the Hartung-Knapp-Sidik-Jonkman method (Hartung & Knapp, 2001; Sidik & Jonkman, 2003) within the Meta package (Schwarzer, 2007).

In order to address the individual research questions, a nested approach to data analysis was taken: Subgroup analyses using random-effects models were performed using the Meta package (Schwarzer, 2007) and effect sizes, 95% CIs,

Z-statistics, *p*-values, *I*<sup>2</sup>, and PIs for subgroups were calculated. Apart from subgroup analysis E, which explored the effect of financial reward on LE (discussed below), all analyses included at least 10 observations, which is considered the minimum for performing this type of analysis (Higgins et al., 2019).

Motivational factors were grouped into five categories: financial reward, perceived competence, feedback, evaluative threat, and individual traits (see Categorization of motivational factors). Outcomes of LE studies were categorized as subjective, behavioral, or physiological (see Measures of LE).

Seven sets of meta-analyses were conducted:

- First, a meta-analysis of all studies was carried out to determine the overall effects of motivational factors on LE outcomes.
- To address RQ1, which asked whether there are differences in LE dependent upon the type of motivational factor, we conducted subgroup analysis A (Table 3) to compare effect sizes of motivational factors on all LE outcomes.
- RQ2 asked whether effect sizes differ depending upon the type of LE outcome measure (i.e., subjective, behavioral, or physiological) and whether certain exemplars of LE outcome measurement categories capture larger effects than others. Subgroup analysis B (Table 5) compared the effect sizes captured using different types of LE outcomes.
- Subgroup analyses C and D (Table 5) compared effect sizes for specific behavioral and physiological measures. An insufficient number of outcomes (i.e., fewer than 10 observations) precluded a subgroup analysis of subjective measures.
- Finally, to address RQ3, which asked which combinations of motivational factors/LE outcomes produced the largest/smallest effects, a series of subgroup analyses were carried out (subgroup analyses E to G) (Table 6), comparing effect sizes for different types of LE outcomes for specific motivational factors.

An alpha level of .05 was used throughout, except for post hoc analyses where a Bonferroni adjustment was made for the number of comparisons. For post hoc analyses, the dmetar package (Harrer et al., 2019) in R (R Core Team, 2018) was used to test for differences in effect sizes between subgroups in a random-effects model, following the procedure of Borenstein and Higgins (2013) to calculate a *Q* statistic and its associated *p*-value. These analyses included only two categories of interest (e.g., individual traits and perceived competence to compare effect sizes for specific motivational factors).

**Table 3.** Meta-Analysis of the Effects of Motivational Factors on all LE Outcome Measurements.

LE outcome	Subgroup	Number of outcomes (contributing studies)	Pooled sample size	Effect size (SMD) 95% CI	Z-statistic <i>p</i> -value	<i>I</i> <sup>2</sup>	Prediction interval
<b>Subgroup analysis A: Comparison of extrinsic manipulations of motivation (k = 184) (Q(3) = 11.91, <i>p</i> = .008) PI = [-.31, 1.00]*</b>							
All	Financial reward	44 (16)	1,046	<b>0.40**</b> [0.28, 0.52]	6.30 (<.001)	67.2 [55.1, 76.0]	[-.29, 1.09]
All	Evaluative threat	23 (3)	574	<b>0.43**</b> [0.27, 0.61]	5.01 (<.001)	47.3 [14.2, 67.6]	[-.18, 1.06]
All	Perceived competence	12 (5)	769	<b>0.67**</b> [0.42, 0.92]	5.22 (<.001)	62.9 [30.9, 80.1]	[-.14, 1.48]
All	Individual traits	105 (18)	5,431	<b>0.27**</b> [0.19, 0.34]	7.09 (<.001)	68.8 [61.9, 74.4]	[-.29, 1.09]

Bold and \**p* < .05, \*\**p* < .001. Note that some studies contribute outcomes to multiple groups. Feedback is not included in this analysis because of insufficient number of outcomes in the included studies.

**Table 4.** Post Hoc Tests for Subgroup Analysis A, Averaged Over all Types of LE Outcome Measurements.

LE outcome	Post hoc test	Q-statistic	p-value
All	Financial reward/Evaluative threat	0.14	.711
All	Financial reward/Perceived competence	3.55	.059
All	Financial reward/Individual traits	2.97	.084
All	Evaluative threat/Perceived competence	2.18	.140
All	Evaluative threat/Individual traits	3.07	.080
All	Perceived competence/Individual traits	8.81	<b>.003*</b>

Bold and \*p < .007 (adjusted alpha level for number of comparisons).

## Results

### Screening Results

Figure 1 shows the PRISMA flowchart for the literature search. The search returned 4,639 studies. After removing duplicates, 3,504 studies remained. After screening titles and abstracts, the full text of 64 studies were assessed for eligibility. An additional 28 studies were sourced by checking the reference list of studies selected for full-text screening and checking citations to these studies. After screening the full text of the total 92 studies, 44 studies were excluded: 19 were excluded because studies did not include an LE outcome, 19 did not manipulate motivation or measure personality traits, 3 did not manipulate motivation or measure LE, 1 was a qualitative study, 1 study used visual rather than auditory stimuli and 1 study combined results for an auditory task with results from a visual task. Forty-eight studies were eligible for inclusion in the review. A summary of the 48 included studies is provided in Table 2. This summary includes sample size, demographic information (ages, hearing status), details of the stimuli used, motivation manipulation/personality trait considered, listening demand manipulations and main findings for each LE outcome.

Of the 48 included studies, only 43 of these studies were suitable for inclusion in the meta-analysis (see Meta-analysis strategy). Four articles (Cahill & Polich, 1992; Cox et al., 1999; Pritchard, 1989; Sewell, 1985) did not provide relevant data to be included with the meta-analysis. A further study (Beatty et al., 1980) was not included in the meta-analysis as this study involved the use of course credit for students as a performance-based reward, rather than financial reward. Included studies often contained multiple outcome measures, leading to a total of 187 outcomes from 43 studies that were included in the meta-analysis.

**Table 5.** Subgroup Analyses Investigating the Influence of LE Outcome Measurement on the Relationship Between Motivational Factors and LE.

LE outcome	Subgroup	Number of outcomes (contributing studies)	Pooled sample size	Effect size (SMD) 95% CI	Z-statistic p-value	I <sup>2</sup>	Prediction interval
<b>Subgroup analysis B: Comparison of LE outcome measurement for all motivational factors (k = 187) (Q(2) = 1.92, p = .383, PI = [-.30, .99])</b>							
Behavioral	-	70 (30)	3,756	<b>0.29**</b> [0.20, 0.39]	5.94 (<.001)	72.4 [65.0, 78.2]	[-.38, 0.97]
Physiological	-	99 (24)	3,536	<b>0.37**</b> [0.29, 0.45]	8.99 (<.001)	63.7 [55.0, 70.7]	[-.26, 1.00]
Subjective	-	18 (6)	600	<b>0.42**</b> [0.20, 0.64]	3.78 (<.001)	70.2 [51.8, 81.6]	[-.42, 1.27]
<b>Subgroup analysis C: Effects of all motivational factors on behavioral LE outcomes (k = 74) by type (Q(1) = 0.01, p = .911, PI = [-.38, .97])</b>							
Behavioral	Performance accuracy	55 (28)	2,969	<b>0.29**</b> [0.18, 0.41]	5.03 (<.001)	73.6 [65.7, 79.7]	[-.43, 1.01]
Behavioral	RTs	15 (10)	787	<b>0.30**</b> [0.13, 0.48]	3.41 (<.001)	65.2 [39.7, 79.9]	[-.30, .91]
<b>Subgroup analysis D: Effects of all motivational factors on physiological LE outcomes (k = 91) by type (Q(1) = 0.11, p = .743, PI = [-.24, .92])</b>							
Physiological	EEG	70 (12)	2,577	<b>0.35**</b> [0.26, 0.44]	7.29 (<.001)	61.6 [50.2, 70.3]	[-.27, .97]
Physiological	Pupillometry	21 (8)	758	<b>0.32**</b> [0.17, 0.47]	4.26 (<.001)	55.5 [27.2, 72.7]	[-.21, .85]

Bold and \*p < .05, \*\* p < .001. Note that some studies contribute outcomes to multiple groups.

**Table 6.** Meta-Analysis of Effects of Combining Different Motivational Factors with Different Types of LE Outcome Measurements.

LE outcome	Subgroup	Number of outcomes (contributing studies)	Pooled sample size	Effect size (SMD) 95% CI	Z-statistic p-value	$I^2$	Prediction interval
<b>Subgroup analysis E: Effects of financial reward on LE (k = 44) by outcome type (Q(2) = 9.53, p = .009, PI = [-.29, 1.09])*</b>							
Behavioral	Financial reward	21 (12)	460	<b>0.34**</b> [0.16, 0.52]	3.66 (<.001)	67.3 [48.5, 79.3]	[-.39, 1.06]
Physiological	Financial reward	17 (9)	430	<b>0.58**</b> [0.37, 0.80]	5.35 (<.001)	72.5 [55.4, 83.1]	[-.24, 1.41]
Subjective	Financial reward	6 (3)	156	<b>0.11*</b> [-0.10, 0.32]	1.05 (0.03)	0 [0, 0]	[-.19, .41]
<b>Subgroup analysis F: Effects of individual traits on LE (k = 103) by outcome type (Q(1) = 5.91, p = .02, PI = [-.34, .89])*</b>							
Behavioral	Individual traits	33 (11)	2,420	<b>0.15*</b> [0.03, 0.27]	2.46 (0.014)	71.1 [58.9, 79.6]	[-.44, .74]
Physiological	Individual traits	70 (13)	2,789	<b>0.34**</b> [0.25, 0.44]	7.05 (<.001)	64.0 [53.5, 72.1]	[-.29, .98]
<b>Subgroup analysis G: Effects of evaluative threat on LE (k = 22) by outcome type (Q(1) = 7.42, p = .006, PI = [-.16, 1.01])*</b>							
Physiological	Evaluative threat	10 (2)	269	<b>0.24*</b> [0.07, 0.38]	2.88 (.004)	0 [0, 34.0]	[-.18, .65]
Subjective	Evaluative threat	10 (2)	222	<b>0.72**</b> [0.43, 1.01]	1.61 (<.001)	55.2 [8.7, 78.0]	[-.15, 1.59]

Bold and \*p < .05, \*\*p < .001. Note that some studies contribute outcomes to multiple groups.

### Risk of Bias Assessment

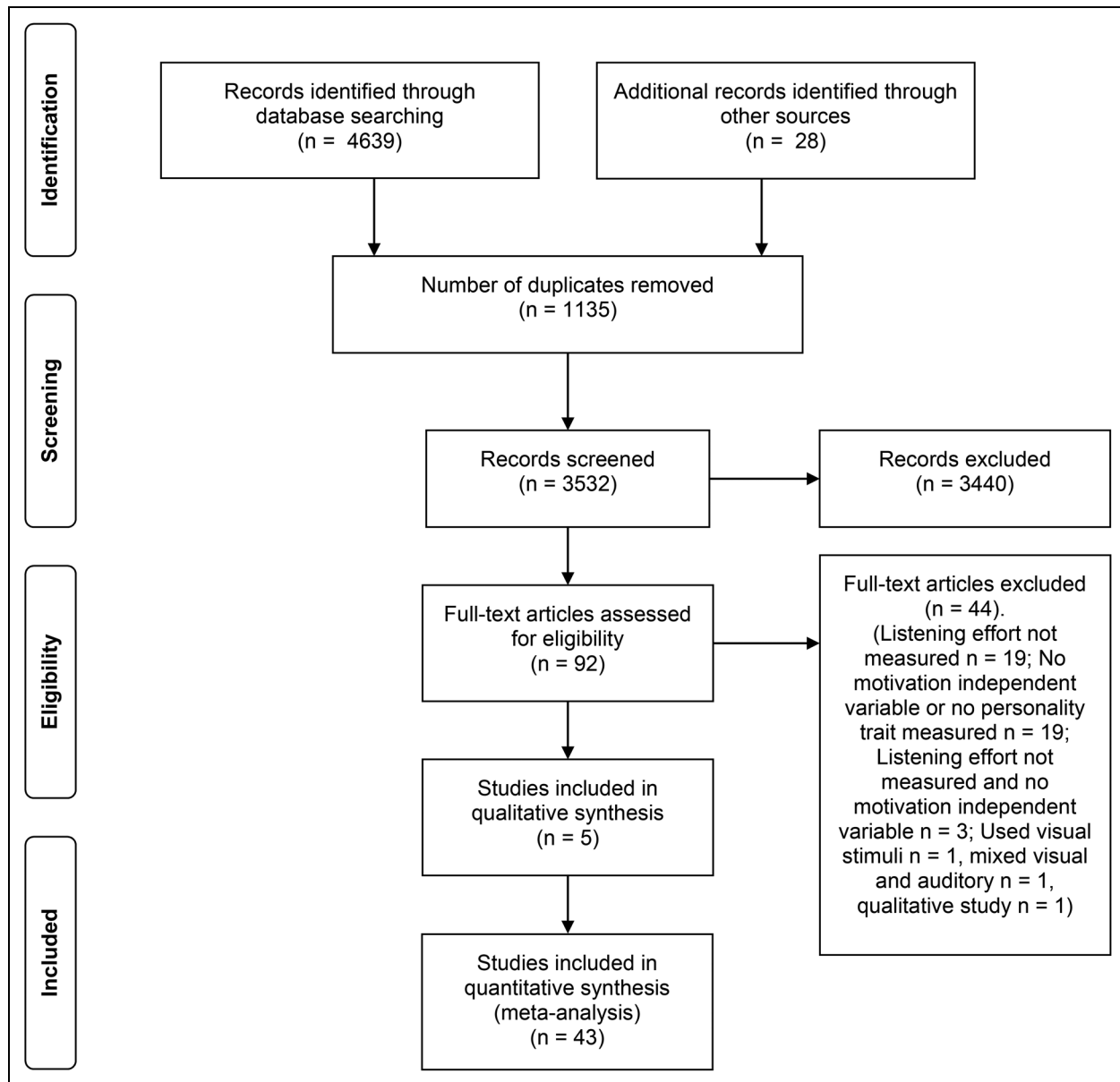
The results of the overall risk of bias assessment are displayed in Figure 2. A breakdown of scores for individual studies is included in the supplemental materials (see Supplementary Table S3). The vast majority of the studies did not include a sample size justification, presenting a high risk of bias. The majority of studies either did not exclude any participants or gave adequate justification for exclusion, presenting a low risk of bias. Reporting of outcomes presented a high risk of bias for ~20% of the studies, and there were some concerns for ~50%. Bias related to outcome reporting may have been present for the following reasons: statistical analyses were not reported for some outcomes, behavioral data were missing from the results, or physiological data were not fully reported. Some studies did not justify conducting analyses on subgroups of participants only. No studies reported conflicts of interest (Q4). The funnel plot of effect sizes (see Supplementary Figure S1) is roughly symmetrical, suggesting no publication bias.

### Qualitative Synthesis

Five studies were not included in the meta-analyses (see Screening results for details). The influence of motivation on LE in these studies was mixed.

Beatty et al. (1980) compared comprehension of a lecture about computer music in students who were, or were not, offered performance-based course credit as a reward. They found significantly greater comprehension for students in the reward compared to no reward group but no significant listening demand/reward group interaction. Sewell (1985) compared listening comprehension of a “scholarly speech” between students who were told the speech would be similar to other classroom lectures (“boring”) or that the speech was on an interesting and relevant topic (“interesting”). Some participants in the interesting group were also offered a financial reward for achieving high levels of listening comprehension (>90%). The effect of group and financial reward was not statistically significant.

Cahill and Polich (1992) compared event-related potential (ERP) components of “extreme groups” of introverts and extraverts during an auditory discrimination task (tones). The listening task contained four difficulty conditions, based upon the probability of the presentation of the target stimulus. The P3 amplitude showed a significant effect of the interaction between personality, probability and gender. No effects were observed for P3 latency. Using a similar auditory discrimination task, Pritchard (1989) found (i) a significant correlation between P3 amplitude and scores on the Lie scale of the Eysenck Personality Questionnaire and (ii) a significant negative correlation between P3 latency and the Neuroticism scale of the Eysenck Personality Questionnaire. These effects were seen in male but not female participants; combined results were not reported.



**Figure 1.** PRISMA flowchart showing the identification, screening, eligibility, and inclusion phases of the literature search.

Cox et al. (1999) investigated personality traits and self-reported ease of communication in adult hearing aid users. Extraversion-introversion and locus of control did not predict ease of communication when participants were aided or unaided. Extraversion-introversion predicted the change scores (i.e., difference between unaided and aided ratings), accounting for 9.5% of the variance.

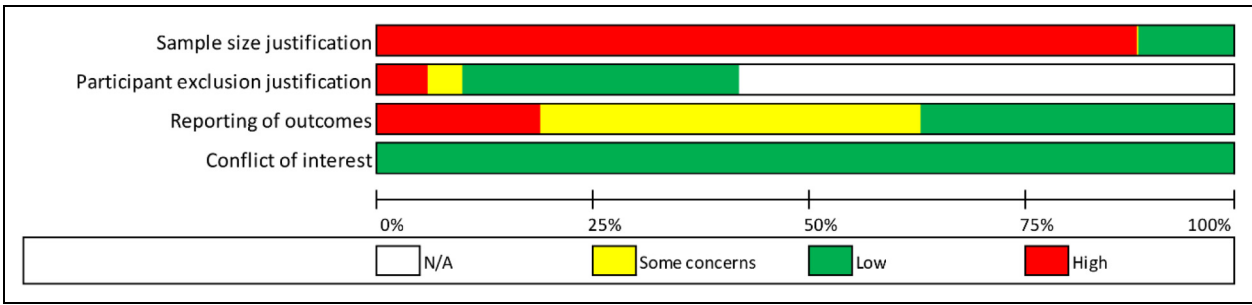
### Meta-Analyses

*Meta-Analysis of Motivational Factors on all LE Outcome Measurements.* A meta-analysis of all studies was conducted to determine the effect of motivational factors (collectively)

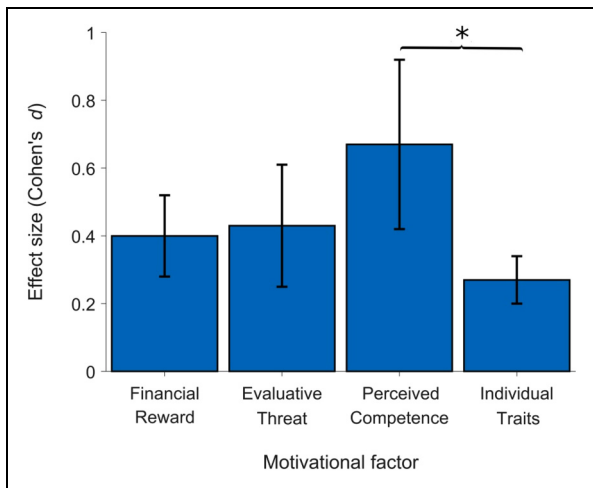
on all LE outcomes. The results revealed an overall positive effect (i.e., LE increased with greater motivation). The random-effects model predicted a small-to-moderate effect size of  $d = .34$  (95% CI = [.28, .40],  $Z = 17.75$ ,  $PI = [-.30, .99]$   $p < .001$ ). Note that heterogeneity was in the moderate-to-high range for the majority of nested analyses (Tables 3–6).

### The Influence of Motivational Factors

Table 3 shows the results of subgroup analysis A, which compared the effects of motivational factors on each LE outcome. The type of motivational factor significantly



**Figure 2.** Summary of risk of bias scores for all included studies. Full details of the scoring questions and descriptions of the response categories are shown in Supplementary Table S1. Q1: Did the authors include a sample size justification? Q2: If any participant data is excluded from the analysis is a clear justification given? Q3: Were all the outcome measures in the methods included in the results? Q4: Were there any conflicts of interest? Overall percentage scores were calculated for each level of bias, with green indicating low bias, yellow some concerns, red high and white not-applicable.



**Figure 3.** Effect sizes (Cohen's *d*) for the influence of specific motivational factors on listening effort. Error bars represent 95% confidence intervals for the effect size. Perceived competence had a significantly larger effect compared to individual traits ( $p = .003$ ).

influenced LE ( $Q(3) = 11.91$ ;  $p = .008$ ,  $PI = [-.33, 1.00]$ ). To investigate the significant effect of motivational factor type, post hoc analyses were carried out (see Table 4 and Figure 3). A significantly smaller effect size for individual traits was found when compared to perceived competence ( $Q(1) = 8.81$ ,  $p = .003$ ). No other post hoc comparisons were significant.

**The Influence of LE Outcome Measurement**

Table 5 shows the results of subgroup analysis B, which investigated the influence of LE outcome type. While a positive effect of motivation on LE was found for every LE outcome, the size of the effect did not differ between outcomes ( $Q(2) = 1.92$ ;  $p = .383$ ,  $PI = [-.33, .99]$ ). The results of subgroup analysis C, which compared effect

**Table 7.** Post Hoc Tests for Subgroup Analysis E.

Motivational factor	Post hoc test	Q-statistic	p-value
Financial reward	Behavioral/ Physiological	2.98	.084
Financial reward	Behavioral/Subjective	2.54	.110
Financial reward	Physiological/Subjective	9.53	<b>.002*</b>

Bold and \* $p < .017$  (adjusted alpha level for number of comparisons).

sizes for specific behavioral measures, are reported in Table 5. Motivation had a similar-sized small effect, regardless of whether studies measured LE as performance accuracy or RTs ( $Q(1) = .01$ ,  $p = .911$ ,  $PI = [-.38, .97]$ ). Table 5 also shows the results of subgroup analysis D, which compared effect sizes for specific physiological outcomes. Again, motivation had a similar-sized small positive effect when LE was measured using EEG or pupillometry ( $Q(1) = .11$ ,  $p = .743$ ,  $PI = [-.24, .92]$ ).

*Nested Analyses for Specific Motivational Factors Comparing Types of LE Outcomes.*

Table 6 shows the results of subgroup analyses used to address RQ3, which asked whether there are differences in effect sizes for different combinations of motivational factors and LE outcomes. Subgroup analysis E (Table 6) compared effect sizes measured with different types of LE outcomes when studies operationalized motivation as financial reward. This analysis showed significant positive effect sizes for studies using behavioral ( $d = .34$ , 95% CI = [.16, .52],  $Z = 3.66$ ,  $p < .001$ ,  $PI = [-.39, 1.06]$ ), physiological ( $d = .58$ , 95% CI = [.37, .80],  $Z = 5.35$ ,  $p < .001$ ,  $PI = [-.24, 1.41]$ ) and subjective measures ( $d = .11$ , 95% CI = [-.10, .32],  $Z = 1.05$ ,  $p = .03$ ,  $PI = [-.19, .41]$ ). When motivation was operationalized as financial reward, there was a significant difference in effect sizes depending upon the type of LE outcome ( $Q(2) = 9.53$ ,  $p = .009$ ,  $PI = [-.29, 1.09]$ ). To investigate the source of this difference, post hoc analyses were carried out to compare effect sizes

measured by each type of LE outcome (Table 7 and Figure 4). A significantly larger effect of financial reward was found for physiological measures of LE when compared to subjective measures ( $Q(1)=9.53, p=.002$ ).

Subgroup analysis F (Table 6) compared effect sizes measured with different types of LE outcomes for studies investigating individual traits. Subjective outcomes were not included in this analysis due to limited numbers of eligible studies. All types of LE outcomes captured positive effects, but the size of the effect depended upon the type of LE outcome ( $Q(1)=5.91, p=.020, PI=[-.34, .89]$ ): a significantly greater effect was seen for physiological ( $d=.34, 95\% CI=[.25, .44], Z=7.05, p<.001, PI=[-.29, .98]$ ) compared to behavioral ( $d=.15, 95\% CI=[.03, .27], Z=2.46, p=.014, PI=[-.44, .74]$ ) outcomes.

Finally, subgroup analysis G (Table 6) considered studies operationalizing motivation as evaluative threat. Due to limited numbers of eligible studies, behavioral outcomes were omitted from this analysis. Again, positive effects were captured using all types of LE outcomes, but the type of LE outcome influenced the size of the effect ( $Q(1)=7.42, p=.006, PI=[-.16, 1.01]$ ): a significantly greater positive effect was seen for subjective ( $d=.72, 95\% CI=[.43, 1.01], Z=1.61, p<.001, PI=[-.15, 1.59]$ ) compared to physiological outcomes ( $d=.24, 95\% CI=[.07, .38], Z=2.88, p=.004, PI=[-.18, .65]$ ).

## Discussion

A systematic review and meta-analysis was carried out to quantify the relationship between motivational factors and types of LE outcomes for the first time. We used random-effects meta-analyses with subgroups to investigate differences in effect sizes for different motivational factors and types of LE outcomes on the expenditure or experience of LE. When LE outcomes were considered collectively, nested analyses showed differences in effect sizes for motivational factors (perceived competence cf. individual traits). When all motivational factors were combined, similar effect sizes were measured using subjective, behavioral and physiological LE outcomes. Comparisons of specific exemplars of LE outcome measurements (where permitted due to study numbers) revealed similar effect sizes. However, for certain motivational factors (financial reward, individual traits and evaluative threat), the type of LE outcome influenced the size of the measured effect.

## Quality of Evidence

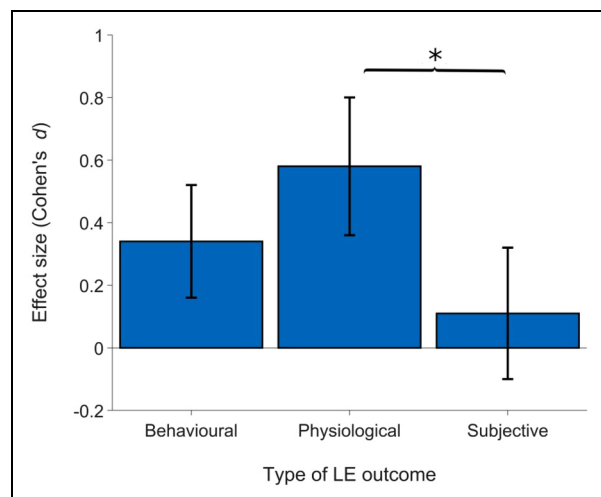
The risk of bias was generally high due to lack of sample size justification. Concerns were also raised about missing or incomplete outcome reporting. These quality issues reduce confidence in the effect sizes presented here.

## The Influence of Motivational Factors on LE

The primary meta-analysis (which considered all motivational factors and all LE outcomes) suggested that listeners expended or experienced greater effort if motivation was high. This supports a role of motivation in the allocation of resources to listening tasks, as outlined by several models and frameworks for example FUEL (Pichora-Fuller et al., 2016), MIT (Brehm & Self, 1989), CET (Kruglanski et al., 2012), and the Model of Listening Engagement (Herrmann & Johnsrude, 2020).

In our meta-analysis, all motivational factors showed positive effects on LE outcomes. This result may have occurred because these motivational factors increase LE investment by appealing to basic psychological needs (Dweck, 2017). Need-press theory (Murray, 1938) predicts that environmental factors, or “presses”, influence how psychological needs regulate behavior. The motivational factors considered in this review may activate different underlying needs, for example financial reward and perceived competence may “press” upon needs in the domains of materialism and human power, respectively, leading to increased LE. Individual traits are also posited to be related to underlying psychological needs, which give rise to characteristic styles of goal pursuit (Dweck, 2017).

Studies operationalizing motivation as individual traits captured significantly smaller effect sizes compared to extrinsic manipulations of motivation, that is, perceived competence. As variability in individual traits within participant samples can be low, some of the studies (Beauducel et al., 2006; Brocke et al., 1996; Cahill & Polich, 1992; Daruna et al., 1985; Ditraglia & Polich, 1991; Enge et al., 2008; Ortiz & Maojo, 1993; Rammsayer & Stahl, 2004; Schmidt et al., 2004; Tune, 1966) compared participants at the



**Figure 4.** Effect sizes (Cohen's  $d$ ) measured using different types of LE outcome measurements when motivation is operationalized as financial reward. Error bars represent 95% confidence intervals for the effect size. Physiological measures indexed significantly larger effects compared to subjective measures of LE ( $p=.002$ ).



extreme ends of the distribution (usually those scoring in the upper and upper lower quartiles). This procedure is commonly used within personality research to increase statistical power (see Preacher et al., 2005 for discussion of the strengths and limitations of this approach). For researchers interested in the connection between motivational trait differences and LE, an extreme group procedure is worth considering to increase the likelihood of producing robust results.

### *The Influence of Outcome Selection on LE*

An overall positive effect of motivational factors was detected using subjective, behavioral and physiological LE outcome measurements, with no significant differences in effect sizes between outcome types. However, there may be differences in sensitivity to motivation for the different types of LE outcomes which are not apparent in the subgroup analyses presented here due to the broad categorization of subjective, behavioral, or physiological LE outcomes. The inclusion of different measures within the same LE outcome domain may contribute to the high heterogeneity in the subgroup analyses (see Caveats).

Finer-grained comparison was limited due to the small numbers of studies, however, we were able to compare effects measured using two behavioral (RT and performance accuracy) and two physiological LE outcomes (pupillometry and EEG). Motivation produced a similar-sized small increase in LE when measured by either performance accuracy or RTs. The FUEL posits that when an individual's resource capacity is under strain or exceeded, performance errors and a reduction in processing speed may occur, which can allow inferences about LE. The FUEL also notes that RTs are not a *process pure measure* (Pichora-Fuller et al., 2016) as they may also index other aspects of cognition such as memory. Improved speech recognition accuracy may be interpreted as an increase in LE (e.g., DeWall et al., 2011; Plain et al., 2020; Zhang et al., 2019). However, performance accuracy may also be considered as a measure of achievement, rather than effort (Gendolla & Richter, 2010). Some studies found that performance accuracy remained stable while other measures of LE changed, for example an increase in RTs but no change in task accuracy in response to greater listening demands (Sarapalis et al., 2009; Strand et al., 2020). The divergence between task accuracy and RTs in these studies supports a theoretical distinction between performance effectiveness (i.e., quality of task performance) and processing efficiency (effectiveness in relation to effort) (Eysenck & Calvo, 1992). In the present meta-analysis, similar sized positive effects were measured with both performance accuracy and RTs, but these outcomes may not necessarily index the same LE construct.

Similar positive effect sizes were also measured using pupillometry and EEG. Studies which contributed towards the overall positive effect for EEG outcomes (e.g., Brocke et al., 1996; Carrillo-de-la-Pena and Cadaveira, 2000; Enge et al., 2008; Ortiz and Maojo, 1993) found an increased P3

amplitude with motivation or motivational traits. For pupillometry, the overall positive effect reflects results showing increased pupil dilation in high compared to low motivation conditions (e.g., Bijleveld et al., 2009; Koelewijn et al., 2018; Zekveld et al., 2019; Zhang et al., 2019). P3 amplitude and pupil dilation may both index activity in the locus coeruleus noradrenergic system, which is associated with regulation of task engagement (Murphy et al., 2011). Due to a limited number of outcomes in the included studies, only these two specific physiological measures could be compared in the present study. Notably, a few studies found physiological effects of motivation (change in pupil size or cardiac response) in the absence of behavioral and/or subjective effects (Koelewijn et al., 2018; Richter, 2016). Increased physiological arousal may reflect the "cost" required to maintain performance under increased task demands (Kahneman, 1973). However, concurrent changes in behavioral and/or subjective measures would increase confidence that the physiological changes reflect the influence of motivation on LE as physiological measures of LE may index multiple processes. For example, Zekveld et al. (2018) discuss the sensitivity of the pupil dilation response to various task manipulations and intra-individual differences.

### *Combinations of Motivational Factors and Types of LE Outcomes Influence LE Effect Sizes*

Subgroup analyses investigated combinations of motivational factors and types of LE outcomes. These analyses revealed significant differences in effect sizes for financial reward, evaluative threat and individual traits, depending upon the type of LE outcome. These differential effects may help inform research planning decisions when investigating motivation and LE. For instance, if researchers are interested in the effect of evaluative threat on LE, using subjective rather than physiological measures would be advisable, based on the results of the meta-analyses presented here. When investigating individual traits or financial reward, physiological measures may be optimal compared with behavioral and subjective measures, respectively. However, note that the effect size associated with subjective measures was based on only six outcomes. At least 10 observations are recommended for subgroup analyses (Higgins et al., 2019). However, 10 observations would be insufficient if covariates are distributed unevenly across studies and, in general, subgroup analyses should be interpreted cautiously where the number of observations is low (Higgins et al., 2019). Although certain combinations of motivational traits/LE outcomes may produce bigger effect sizes than others, it is important to note that PIs encompassed zero for all factor/outcome type combinations and therefore null or negative effects may be observed in future studies.

Differences in effect sizes are consistent with the notion that LE is a multi-factorial construct (Alhanbali et al.,

2019; Herrmann & Johnsrude, 2020; Hughes et al., 2018; McMahon et al., 2016; Pichora-Fuller et al., 2016; Strand et al., 2018; Strauss & Francis, 2017). However, here we quantified differences between types of LE outcomes only when considering specific motivational factors. Particular types of LE outcomes may be sensitive to certain motivational factors for several reasons. For example, biological differences may underlie trait behaviors, for example relative to introverts, extraverts are suggested to have lower levels of cortical arousal (Eysenck, 1967) and this may explain why larger effect sizes were measured for individual traits using physiological outcomes. The finding of larger effect sizes for subjective compared to physiological LE outcomes, when motivation is operationalized as evaluative threat, may be due to the small number of studies that used cardiac reactivity to measure LE: cardiac reactivity is posited to be particularly sensitive to increases in effort due to fear of social evaluation (Gendolla & Richter, 2010). The large effect size for subjective LE outcomes suggests that evaluative threat elicits a range of effort-related listening experiences, as the studies included in the meta-analysis used questions gauging effort, likelihood of quitting and desire to control the situation (Picou & Ricketts, 2014; Zekveld et al., 2019). However, evaluative threats may increase anxious thoughts which can occupy working memory (Eysenck et al., 2007; Putwain et al., 2014) and effortful listening may rely on working memory capacity (Rönnerberg et al., 2019). High levels of anxiety may lead to reduced processing capacity for the task at hand and result in poorer performance (Eysenck et al., 2007; Eysenck & Calvo, 1992).

### Caveats

Heterogeneity was in the moderate-to-high range for the subgroup analyses included in this systematic review. Inconsistency measures of 25%, 50%, and 75% are usually interpreted as low, moderate and high levels of heterogeneity, respectively (Higgins & Thompson, 2002). The high levels of heterogeneity in the meta-analyses presented here are comparable with other systematic reviews of experimental psychology research (e.g., Hirst et al., 2018; Rung and Madden, 2018). Within audiology research, a meta-analysis of experimental studies investigating whether probe-tube verification of real-ear hearing aid amplification characteristics improves outcomes in adults identified high heterogeneity ( $I^2$ : 61–100%) for a number of subgroup analyses (Almufarrij et al., 2021). Heterogeneity in the present analysis may have arisen due to between-study differences in participant characteristics (age, gender, hearing level, etc.), variation in the operationalization of listening demands (e.g., single or dual tasks, vocoded speech or speech presented in a noise background) and stimulus complexity (e.g., tones, digits, or spoken sentences). There is also variation in how motivational factors were operationalized within each category. For instance, studies using financial reward to motivate participants offered different amounts of

financial reward, for example in the high motivation condition Bijleveld et al. (2009) offered 50 Euros, Koelewijn et al. (2018) offered 5 Euros. Some studies compared effects of “high” and “low” levels of financial reward (Asbjörnsen et al., 1992; Bijleveld et al., 2009; Carolan et al., 2021; Kahneman et al., 1968; Kahneman & Peavler, 1969), other designs compared a financial reward condition to a no financial reward condition (Kang et al., 2017; Snyder & Snyder, 1956) or required strategic allocation of LE to maximize reward (Gilzenrat et al., 2010; Zhang, 2017; Zhang et al., 2019).

As the practical implications of the observed heterogeneity, expressed as  $I^2$ , are not always clear, we also calculated PIs, which show the range of effect sizes that researchers can expect when planning future studies using similar paradigms to those included in the meta-analysis. Where there is substantial heterogeneity in a meta-analysis, the prediction interval has a wider range than the 95% CI of the standardized mean difference; thus it may be misleading to only present the standardized mean difference and its CI, since the PIs may reveal potential values on both sides of the null (IntHout et al., 2016). All the PIs in the present meta-analysis contained values on both sides of the null, thus inconsistent results are possible using the same motivational factor/LE outcome. For example, this could explain why Richter (2016) found a significant increase in LE measured using cardiac reactivity (pre-ejection period, PEP) when participants were offered performance-based financial reward, while Plain et al. (2020), found null results for the effects of financial reward on PEP reactivity.

As discussed above, the majority of studies in this review did not include a power analysis. Regardless of the size of a statistical effect, the occurrence of both type S and type M errors is substantially inflated when statistical power is low (Gelman & Carlin, 2014). A type S error is the probability that the estimated effect has the wrong sign, that is that the estimate indicates a positive effect when in fact the true effect is negative or *vice versa* (Kirby & Sonderegger, 2018). Type S errors may explain inconsistencies within the literature for some measures, for example these errors may account for the differential effects of motivation on RTs reported in the included studies (Carolan et al., 2021; Carrillo-de-la-Pena & Cadaveira, 2000; Kang et al., 2017; Strand et al., 2018; Weis et al., 2013). A type M error is the extent to which the magnitude of an effect of interest is exaggerated, based on the estimated effect size of the study. It is possible that type M errors may account for outliers in the meta-analyses (see Supplementary Figure S1). We therefore urge researchers in the field to ensure that future studies of the effects of motivation on LE are adequately powered and that the results of power calculations are fully reported.

The effect of motivation is proposed to be greatest when tasks are demanding (Brehm & Self, 1989). If listening is perceived to be too difficult then participants may decide to disengage, which may have impacted upon the results of the meta-analyses presented here: we decided to include either

an effect size for the interaction, or if not available, the effect size from the most demanding conditions, when studies utilized multiple listening demand conditions (see Methods: Meta-analysis strategy).

## Conclusion

All motivational factors considered here, in the first meta-analysis of motivation and LE, increased the experience and/or exertion of LE. Therefore it is important to consider motivational context when designing and interpreting LE studies. External manipulations of motivation (e.g., perceived competence) elicited more robust effect sizes compared to individual traits. For certain motivational factors (individual traits, evaluative threat, financial rewards), effect sizes were dependent upon LE outcome measurement. Further research into the relationship between motivation and LE should be guided by this review, which, for the first time, highlights combinations of LE outcomes and motivational factors that are likely to produce the most robust effects on LE.

## Acknowledgments

This work was supported by the NIHR Manchester Biomedical Research Centre (BRC-1215-20007). The authors thank the corresponding authors of the studies for sharing data not reported in the publications and for responding to our queries.





## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

This work was funded by the NIHR Manchester Biomedical Research Centre (BRC-1215-20007).

## ORCID iDs

Peter J Carolan  <https://orcid.org/0000-0003-4386-0310>  
 Antje Heinrich  <http://orcid.org/0000-0002-5940-622X>  
 Kevin J Munro  <https://orcid.org/0000-0001-6543-9098>  
 Rebecca E. Millman  <http://orcid.org/0000-0001-8606-0167>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Alhanbali S., Dawes P., Lloyd S., & Munro K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear and Hearing, 38*(1), 39–48. <https://doi.org/10.1097/Aud.0000000000000361>
- Alhanbali S., Dawes P., Lloyd S., & Munro K. J. (2018). Hearing handicap and speech recognition correlate With self-reported listening effort and fatigue. *Ear and Hearing, 39*(3), 470–474. <https://doi.org/10.1097/Aud.0000000000000515>
- Alhanbali S., Dawes P., Millman R. E., & Munro K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing, 40*(5), 1084–1097. <https://doi.org/10.1097/AUD.0000000000000697>
- Almufarrij I., Dillon H., & Munro K. J. (2021). Does probe-tube verification of real-Ear hearing Aid amplification characteristics improve outcomes in adults? A systematic review and meta-analysis. *Trends in Hearing, 25*, 2331216521999563. <https://doi.org/10.1177/2331216521999563>
- Asbjörnsen A., Hugdahl K., & Bryden M. P. (1992). Manipulations of subjects' level of arousal in dichotic listening. *Brain and Cognition, 19*(2), 183–194. [https://doi.org/10.1016/0278-2626\(92\)90044-M](https://doi.org/10.1016/0278-2626(92)90044-M)
- Bakan P. (1959). Extraversion-Introversion and improvement in an auditory vigilance task. *British Journal of Psychology, 50*(4), 325–332. <https://doi.org/10.1111/j.2044-8295.1959.tb00711.x>
- Beatty M. J., Behnke R. R., & Froelich D. L. (1980). Effects of achievement incentive and presentation rate on listening comprehension. *Quarterly Journal of Speech, 66*(2), 193–200. <https://doi.org/10.1080/00335638009383515>
- Beauducel A., Brocke B., & Leue A. (2006). Energetical bases of extraversion: Effort, arousal, EEG, and performance. *International Journal of Psychophysiology, 62*(2), 212–223. <https://doi.org/10.1016/j.ijpsycho.2005.12.001>
- Bijleveld E., Custers R., & Aarts H. (2009). The unconscious eye opener: Pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychological Science, 20*(11), 1313–1315. <https://doi.org/10.1111/j.1467-9280.2009.02443.x>
- Borenstein M., & Higgins J. P. T. (2013). Meta-Analysis and subgroups. *Prevention Science, 14*(2), 134–143. <https://doi.org/10.1007/s11121-013-0377-7>
- Brehm J. W., & Self E. A. (1989). The intensity of motivation. *Annual Review of Psychology, 40*, 109–131. <https://doi.org/10.1146/annurev.ps.40.020189.000545>
- Brocke B., Tasche K. G., & Beauducel A. (1996). Biopsychological foundations of extraversion: Differential effort reactivity and the differential P300 effect. *Personality and Individual Differences, 21*(5), 727–738. [https://doi.org/10.1016/0191-8869\(96\)00117-1](https://doi.org/10.1016/0191-8869(96)00117-1)
- Buchanan E., Scofield J., & Valentine K. D. (2017). MOTE: Effect Size and Confidence Interval Calculator (Version 1.0.2).
- Burgers C., Eden A., Engelenburg M., & Buningh S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior, 48*, 94–103. <https://doi.org/10.1016/j.chb.2015.01.038>
- Cacioppo J. T., & Petty R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cahill J. M., & Polich J. (1992). P300, probability, and introverted/extroverted personality types. *Biological Psychology, 33*(1), 23–35. [https://doi.org/10.1016/0301-0511\(92\)90003-D](https://doi.org/10.1016/0301-0511(92)90003-D)
- Carolan P. J., Heinrich A., Munro K. J., & Millman R. E. (2021). Financial reward has differential effects on behavioural and self-report measures of listening effort. *International Journal of Audiology, 1–11*. <https://doi.org/10.1080/14992027.2021.1884907>
- Carrillo-de-la-Pena M. T., & Cadaveira F. (2000). The effect of motivational instructions on P300 amplitude. *Clinical Neurophysiology, 30*(4), 232–239. [https://doi.org/10.1016/s0987-7053\(00\)00220-3](https://doi.org/10.1016/s0987-7053(00)00220-3)

- Centre for Research and Dissemination (2009). *Systematic reviews: CRD's guidance for undertaking reviews in healthcare*. Centre for Reviews & Dissemination.
- Costa P. T., & McCrae R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO personality inventory. *Journal of Personality and Social Psychology*, 54(5), 853–863. <https://doi.org/10.1037/0022-3514.54.5.853>
- Cox R. M., Alexander G. C., & Gray G. (1999). Personality and the subjective assessment of hearing aids. *Journal of the American Academy of Audiology*, 10(1), 1–13.
- Cumming G. (2012). *Understanding the New statistics: Effect sizes, confidence intervals and meta-analysis*. Routledge.
- Cumming G., & Calin-Jageman R. (2017). *Introduction to the New statistics: Estimation, open science, and beyond*. Routledge.
- Cusella L. P. (1982). The effects of source expertise and feedback valence on intrinsic motivation. *Human Communication Research*, 9(1), 17–32. <https://doi.org/10.1111/j.1468-2958.1982.tb00680.x>
- Daruna J. H., Karrer R., & Rosen A. J. (1985). Introversion, attention and the late positive component of event-related potentials. *Biological Psychology*, 20(4), 249–259. [https://doi.org/10.1016/0301-0511\(85\)90001-8](https://doi.org/10.1016/0301-0511(85)90001-8)
- Deci E., Koestner R., & Ryan R. (1999). A meta-analytic review of experiments examining the effect of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–692. <https://doi.org/10.1037/0033-2909.125.6.627>
- DeWall C. N., Baumeister R. F., Mead N. L., & Vohs K. D. (2011). How leaders self-regulate their task performance: Evidence that power promotes diligence, depletion, and disdain. *Journal of Personality and Social Psychology*, 100(1), 47–65. <https://doi.org/10.1037/a0020932>
- Ditraglia G. M., & Polich J. (1991). P300 and introverted/extraverted personality types. *Psychophysiology*, 28(2), 177–184. <https://doi.org/10.1111/j.1469-8986.1991.tb00410.x>
- Doucet C., & Stelmack R. M. (2000). An event-related potential analysis of extraversion and individual differences in cognitive processing speed and response execution. *Journal of Personality and Social Psychology*, 78(5), 956–964. <https://doi.org/10.1037/0022-3514.78.5.956>
- Dryden A., Allen H. A., Henshaw H., & Heinrich A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends in Hearing*, 21, 1–21. <https://doi.org/10.1177/2331216517744675>
- Dweck C. (2017). From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychological Review*, 124(6), 689–719. <https://doi.org/10.1037/rev0000082>
- Elliot A. J. (2006). The hierarchical model of approach-avoidance motivation. *Motivation and Emotion*, 30(2), 111–116. <https://doi.org/10.1007/s11031-006-9028-7>
- Engel S., Fleischhauer M., Brocke B., & Strobel A. (2008). Neurophysiological measures of involuntary and voluntary attention allocation and dispositional differences in need for cognition. *Personality & Social Psychology Bulletin*, 34(6), 862–874. <https://doi.org/10.1177/0146167208315556>
- Eysenck H. J. (1967). *The biological basis of personality*. Thomas.
- Eysenck H. J. (1990). Biological dimensions of personality. In *Handbook of personality: Theory and research* (pp. 244–276). The Guilford Press.
- Eysenck M. W., & Calvo M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion*, 6(6), 409–434. <https://doi.org/10.1080/02699939208409696>
- Eysenck M. W., & Derakshan N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences*, 50(7), 955–960. <https://doi.org/10.1016/j.paid.2010.08.019>
- Eysenck M. W., Derakshan N., Santos R., & Calvo M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion (Washington, D C)*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Francis A. L., Bent T., Schumaker J., Love J., & Silbert N. (2021). Listener characteristics differentially affect self-reported and physiological measures of effort associated with two challenging listening conditions. *Attention, Perception, & Psychophysics*, 83(4), 1818–1841. <https://doi.org/10.3758/s13414-020-02195-9>
- Gale A., Bull R., Penfold V., Colestt M., & Barraclough R. (1972). Extroversion, time of day, vigilance performance, and physiological arousal: Failure to replicate traditional findings. *Psychonomic Science*, 29(1), 1–5. <https://doi.org/10.3758/BF03336547>
- Gelman A., & Carlin J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, 9(6), 641–651. <https://doi.org/10.1177/1745691614551642>
- Gendolla G. H., & Richter M. (2010). Effort mobilization when the self is involved: Some lessons from the cardiovascular system. *Review of General Psychology*, 14(3), 212–226. <https://doi.org/10.1037/a0019742>
- Gilzenrat M. S., Nieuwenhuis S., Jepma M., & Cohen J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioural Neuroscience*, 10(2), 252–269. <https://doi.org/10.3758/cabn.10.2.252>
- Gurrera R. J., O'Donnell B. F., Nestor P. G., Gainski J., & McCarley R. W. (2001). The P3 auditory event-related brain potential indexes major personality traits. *Biological Psychiatry*, 49(11), 922–929. [https://doi.org/10.1016/S0006-3223\(00\)01067-2](https://doi.org/10.1016/S0006-3223(00)01067-2)
- Gurrera R. J., Salisbury D. F., O'Donnell B. F., Nestor P. G., & McCarley R. W. (2005). Auditory P3 indexes personality traits and cognitive function in healthy men and women. *Psychiatry Research*, 133(2), 215–228. <https://doi.org/10.1016/j.psychres.2004.09.009>
- Hackley S. A., & Valle-Inclán F. (1998). Automatic alerting does not speed late motoric processes in a reaction-time task. *Nature*, 391(6669), 786–788. <https://doi.org/10.1038/35849>
- Harrer M., Cuijpers P., & Ebert D. (2019). *Doing Meta-Analysis in R (Version 1.0.0)*.
- Hartung J., & Knapp G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, 20(24), 3875–3889. <https://doi.org/10.1002/sim.1009>
- Henderlong J., & Lepper M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin*, 128(5), 774–795. <https://doi.org/10.1037/0033-2909.128.5.774>

- Herrmann B., & Johnsrude I. S. (2020). A model of listening engagement (MoLE). *Hearing Research*, *397*, 1–12. <https://doi.org/10.1016/j.heares.2020.108016>
- Higgins J. P. T., Altman D. G., Gøtzsche P. C., Jüni P., Moher D., Oxman A. D., Savović J., Schulz K. F., Weeks L., & Sterne J. A. C. (2011). The cochrane collaboration's tool for assessing risk of bias in randomised trials. *Bmj*, *343*, d5928. <https://doi.org/10.1136/bmj.d5928>
- Higgins J. P. T., & Thompson S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. <https://doi.org/10.1002/sim.1186>
- Higgins J., Thomas J., Chandler J., Cumpston M., Li T., Page M., & Welch V. (eds) (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. Retrieved from [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook)
- Hirst R. J., Cragg L., & Allen H. A. (2018). Vision dominates audition in adults but not children: A meta-analysis of the colavita effect. *Neuroscience & Biobehavioral Reviews*, *94*, 286–301. <https://doi.org/10.1016/j.neubiorev.2018.07.012>
- Hodgetts W. E., Aalto D., Ostevik A., & Cummine J. (2018). Changing hearing performance and sound preference with words and expectations: Meaning responses in audiology. *Ear and Hearing*, *40*(3), 615–620. <https://doi.org/10.1097/aud.0000000000000634>
- Horne J. A., & Pettitt A. N. (1985). High incentive effects on vigilance performance during 72 h of total sleep deprivation. *Acta Psychol (Amst)*, *58*(2), 123–139. [https://doi.org/10.1016/0001-6918\(85\)90003-4](https://doi.org/10.1016/0001-6918(85)90003-4)
- Hornsby B. W. Y. (2013). The effects of hearing Aid Use on listening effort and mental fatigue associated With sustained speech processing demands. *Ear and Hearing*, *34*(5), 523–534. <https://doi.org/10.1097/AUD.0b013e31828003d8>
- Hughes S. E., Hutchings H. A., Rapport F. L., McMahon C. M., & Boisvert I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: A grounded theory to establish content validity for a New patient-reported outcome measure. *Ear and Hearing*, *39*(5), 922–934. <https://doi.org/10.1097/AUD.0000000000000553>
- Hullett C., & Levine T. (2003). The overestimation of effect sizes from F values in meta-analysis: The cause and a solution. *Communication Monographs*, *70*(1), 52–67. <https://doi.org/10.1080/715114664>
- IntHout J., Ioannidis J. P. A., Rovers M. M., & Goeman J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ open*, *6*(7), 1–6. <https://doi.org/10.1136/bmjopen-2015-010247>
- John O. P., Naumann L. P., & Soto C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). The Guilford Press.
- Kahneman D. (1973). *Attention and effort*. Englewood Cliffs. Prentice-Hall, Inc.
- Kahneman D., & Beatty J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583–1585. <https://doi.org/10.1126/science.154.3756.1583>
- Kahneman D., & Peavler W. S. (1969). Incentive effects and pupillary changes in association learning. *Journal of Experimental Psychology*, *79*(2), 312–318. <https://doi.org/10.1037/h0026912>
- Kahneman D., Peavler W. S., & Onuska L. (1968). Effects of verbalization and incentive on the pupil response to mental activity. *Canadian Journal of Psychology*, *22*(3), 186–196. <https://doi.org/10.1037/h0082759>
- Kang G., Wang L., & Zhou X. (2017). Reward interacts with modality shift to reduce cross-modal conflict. *Journal of Vision*, *17*(1), 19. <https://doi.org/10.1167/17.1.19>
- Kirby J., & Sonderegger M. (2018). Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics*, *70*, 70–85. <https://doi.org/10.1016/j.wocn.2018.05.005>
- Kluger A. N., & DeNisi A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Koelwijn T., Zekveld A. A., Lunner T., & Kramer S. E. (2018). The effect of reward on listening effort as reflected by the pupil dilation response. *Hearing Research*, *367*, 106–112. <https://doi.org/10.1016/j.heares.2018.07.011>
- Koelwijn T., Zekveld A. A., Lunner T., & Kramer S. E. (2021). The effect of monetary reward on listening effort and sentence recognition. *Hearing Research*, *406*, 108255. <https://doi.org/10.1016/j.heares.2021.108255>
- Koenig S., Uengoer M., & Lachnit H. (2017). Attentional bias for uncertain cues of shock in human fear conditioning: Evidence for attentional learning theory. *Frontiers in Human Neuroscience*, *11*, 266. <https://doi.org/10.3389/fnhum.2017.00266>
- Kok A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577. <https://doi.org/10.1017/S0048577201990559>
- Krenn B., Würth S., & Hergovich A. (2013). The impact of feedback on goal setting and task performance. *Swiss Journal of Psychology*, *72*(2), 79–89. <https://doi.org/10.1024/1421-0185/a000101>
- Kruglanski A. W., Belanger J. J., Chen X., Kopetz C., Pierro A., & Mannetti L. (2012). The energetics of motivated cognition: A force-field analysis. *Psychological Review*, *119*(1), 1–20. <https://doi.org/10.1037/a0025488>
- Krupski A., Raskin D. C., & Bakan P. (1971). Physiological and personality correlates of commission errors in an auditory vigilance task. *Psychophysiology*, *8*(3), 304–311. <https://doi.org/10.1111/j.1469-8986.1971.tb00460.x>
- Lakens D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(863), 1–12. <https://doi.org/10.3389/fpsyg.2013.00863>
- Larisch R., Kotter R., Kehren F., Tosch M., Shah N. J., Kalveram K. T., Jancke L., & Müller-Gärtner H. W. (1999). Motivation effects in a dichotic listening task as evident from functional magnetic resonance imaging in human subjects. *Neuroscience Letters*, *267*(1), 29–32. [https://doi.org/10.1016/s0304-3940\(99\)00312-2](https://doi.org/10.1016/s0304-3940(99)00312-2)
- Lee K., & Ashton M. C. (2012). *The H Factor of Personality: Why Some People are Manipulative, Self-entitled, Materialistic, and Exploitive—and why it Matters for Everyone*. Wilfrid Laurier University Press. ISBN: 9781554588343. *Personality and Individual Differences*. <https://doi.org/10.1016/j.paid.2013.06.010>
- Lenhard W., & Lenhard A. (2016). Calculation of Effect Sizes. Retrieved from [https://www.psychometrica.de/effect\\_size.html](https://www.psychometrica.de/effect_size.html)
- Marslen-Wilson W., & Tyler L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71. [https://doi.org/10.1016/0010-0277\(80\)90015-3](https://doi.org/10.1016/0010-0277(80)90015-3)

- Matthen M. (2016). Effort and displeasure in people who are hard of hearing. *Ear and Hearing, 37*, 28S–34S. <https://doi.org/10.1097/Aud.0000000000000292>
- Mattys S. L., Davis M. H., Bradlow A. R., & Scott S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953–978. <https://doi.org/10.1080/01690965.2012.705006>
- McAuley J. D., Henry M. J., Wedd A., Pleskac T. J., & Cesario J. (2012). Effects of musicality and motivational orientation on auditory category learning: A test of a regulatory-fit hypothesis. *Memory & Cognition, 40*(2), 231–251. <https://doi.org/10.3758/s13421-011-0146-4>
- McCabe K., & Fleeson W. (2012). What is extraversion for? Integrating trait and motivational perspectives and identifying the purpose of extraversion. *Psychological Science, 23*(12), 1498–1505. <https://doi.org/10.1177/0956797612444904>
- McCabe K. O., & Fleeson W. (2016). Are traits useful? Explaining trait manifestations as tools in the pursuit of goals. *Journal of Personality and Social Psychology, 110*(2), 287–301. <https://doi.org/10.1037/a0039490>
- McMahon C. M., Boisvert I., de Lissa P., Granger L., Ibrahim R., Lo C. Y., Miles K., & Graham P. L. (2016). Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Frontiers in Psychology, 7*, 745. [10.3389/fpsyg.2016.00745](https://doi.org/10.3389/fpsyg.2016.00745)
- Miles K., McMahon C., Boisvert I., Ibrahim R., de Lissa P., Graham P., & Lyxell B. (2017). Objective assessment of listening effort: Coregistration of pupillometry and EEG. *Trends in Hearing, 21*, 1–13. <https://doi.org/10.1177/2331216517706396>
- Mirkovic B., Debener S., Schmidt J., Jaeger M., & Neher T. (2019). Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ? *Hearing Research, 377*, 260–270. <https://doi.org/10.1016/j.heares.2019.04.005>
- Moher D., Liberati A., Tetzlaff J., Altman D. G., & Group P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Mumm J., & Mutlu B. (2011). Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior, 27*(5), 1643–1650. <https://doi.org/10.1016/j.chb.2011.02.002>
- Murphy P. R., Robertson I. H., Balsters J. H., & O'Connell R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology, 48*(11), 1532–1543. <https://doi.org/10.1111/j.1469-8986.2011.01226.x>
- Murray H. A. (1938). *Explorations in personality: A clinical and experimental study of fifty Men of college Age*. Oxford University Press.
- Nicholls J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*(3), 328–346. <https://doi.org/10.1037/0033-295X.91.3.328>
- Obleser J., & Weisz N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex, 22*(11), 2466–2477. <https://doi.org/10.1093/cercor/bhr325>
- Okusa M., Shiraishi T., Kubo T., & Nageishi Y. (1999). Effects of discrimination difficulty on cognitive event-related brain potentials in patients with cochlear implants. *Otolaryngology and Head and Neck Surgery, 121*(5), 610–615. [https://doi.org/10.1016/s0194-5998\(99\)70067-7](https://doi.org/10.1016/s0194-5998(99)70067-7)
- Ortiz T., & Maojo V. (1993). Comparison of the P300 wave in introverts and extraverts. *Personality and Individual Differences, 15*(1), 109–112. [https://doi.org/10.1016/0191-8869\(93\)90051-4](https://doi.org/10.1016/0191-8869(93)90051-4)
- Petersen E. B., Wöstmann M., Obleser J., Stenfelt S., & Lunner T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Frontiers in Psychology, 6*, 177–177. <https://doi.org/10.3389/fpsyg.2015.00177>
- Petrie C. R., & Carrel S. D. (1976). The relationship of motivation, listening capability, initial information, and verbal organizational ability to lecture comprehension and retention. *Communication Monographs, 43*(3), 187–194. <https://doi.org/10.1080/03637757609375931>
- Pichora-Fuller M. K., Kramer S. E., Eckert M. A., Edwards B., Hornsby B. W., Humes L. E., Lemke U., Lunner T., Matthen M., Mackersie C. L., Naylor G., Phillips N. A., Richter M., Rudner M., Sommers M. S., Tremblay K. L., & Wingfield A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37*(Suppl 1), 5S–27S. <https://doi.org/10.1097/AUD.0000000000000312>
- Picou E. M., & Ricketts T. A. (2014). Increasing motivation changes subjective reports of listening effort and choice of coping strategy. *International Journal of Audiology, 53*(6), 418–426. <https://doi.org/10.3109/14992027.2014.880814>
- Pisoni D. B., & Tash J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*(2), 285–290. <https://doi.org/10.3758/BF03213946>
- Plain B., Richter M., Zekveld A. A., Lunner T., Bhuiyan T., & Kramer S. E. (2020). Investigating the influences of task demand and reward on cardiac Pre-ejection period reactivity during a speech-in-noise task. *Ear and Hearing, 42*(3), 718–731. <https://doi.org/10.1097/aud.0000000000000971>
- Polich J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 118*(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- Polich J., & Martin S. (1992). P300, cognitive capability, and personality: A correlational study of university undergraduates. *Personality and Individual Differences, 13*(5), 533–543. [https://doi.org/10.1016/0191-8869\(92\)90194-T](https://doi.org/10.1016/0191-8869(92)90194-T)
- Preacher K., Rucker D., MacCallum R., & Nicewander A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10*(2), 178–192. <https://doi.org/10.1037/1082-989X.10.2.178>
- Pritchard W. S. (1989). P300 and EPQ/STPI personality traits. *Personality and Individual Differences, 10*(1), 15–24. [https://doi.org/10.1016/0191-8869\(89\)90172-4](https://doi.org/10.1016/0191-8869(89)90172-4)
- Putwain D. W., Shah J., & Lewis R. (2014). Performance-evaluation threat does not adversely affect verbal working memory in high test-anxious persons. *Journal of Cognitive Education and Psychology, 13*(1), 120–136. <https://doi.org/10.1891/1945-8959.13.1.120>
- R Core Team. (2018). The R Foundation for Statistical Computing Platform. Retrieved from <http://www.R-project.org>

- Ramsayer T., & Stahl J. (2004). Extraversion-related differences in response organization: Evidence from lateralized readiness potentials. *Biological Psychology*, *66*(1), 35–49. <https://doi.org/10.1016/j.biopsycho.2003.08.003>
- Richter M. (2016). The moderating effect of success importance on the relationship between listening demand and listening effort. *Ear and Hearing*, *37*(Suppl 1), 111S–117S. <https://doi.org/10.1097/AUD.0000000000000295>
- Rohatgi A. (2020). WebPlotDigitizer (Version 4.3). Retrieved from <https://automeris.io/WebPlotDigitizer>
- Rosenthal R. (1991). *Meta-Analytic procedures for social research*. Sage Publications, Inc.
- Rönnberg J., Holmer E., & Rudner M. (2019). Cognitive hearing science and ease of language understanding. *International Journal of Audiology*, *58*(5), 247–261. <https://doi.org/10.1080/14992027.2018.1551631>
- Rung J. M., & Madden G. J. (2018). Experimental reductions of delay discounting and impulsive choice: A systematic review and meta-analysis. *Journal of Experimental Psychology; General*, *147*(9), 1349–1381. <https://doi.org/10.1037/xge0000462>
- Sarampalis A., Kalluri S., Edwards B., & Hafter E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230–1240. [https://doi.org/10.1044/1092-4388\(2009/08-0111\)](https://doi.org/10.1044/1092-4388(2009/08-0111))
- Sauseng P., & Klimesch W. (2008). What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience & Biobehavioral Reviews*, *32*(5), 1001–1013. <https://doi.org/10.1016/j.neubiorev.2008.03.014>
- Schmidt A., Beauducel A., Brocke B., & Strobel A. (2004). Vigilance performance and extraversion reconsidered: Some performance differences can indeed be induced. *Personality and Individual Differences*, *36*(6), 1343–1351. [https://doi.org/10.1016/S0191-8869\(03\)00221-6](https://doi.org/10.1016/S0191-8869(03)00221-6)
- Schwarzer G. (2007). Meta: An R package for meta-analysis. *R news*, *7*, 40–45. [https://cran.r-project.org/doc/Rnews/Rnews\\_2007-3.pdf](https://cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf)
- Sewell E. H. (1985). Effects of orienting instructions, monetary incentives, and sex of listeners on listening comprehension. *Perceptual and Motor Skills*, *60*(2), 511–514. <https://doi.org/10.2466/pms.1985.60.2.511>
- Sidik K., & Jonkman J. N. (2003). On constructing confidence intervals for a standardized mean difference in meta-analysis. *Communications in Statistics - Simulation and Computation*, *32*(4), 1191–1203. <https://doi.org/10.1081/SAC-120023885>
- Siegle G., Steinhauer S., Stenger V. A., Konecky R., & Carter C. S. (2003). Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *NeuroImage*, *20*(1), 114–124. [https://doi.org/10.1016/S1053-8119\(03\)00298-2](https://doi.org/10.1016/S1053-8119(03)00298-2)
- Smillie L., Yeo G., Furnham A., & Jackson C. (2006). Benefits of all work and no play: The relationship between neuroticism and performance as a function of resource allocation. *The Journal of Applied Psychology*, *91*(1), 139–155. <https://doi.org/10.1037/0021-9010.91.1.139>
- Snyder F. W., & Snyder C. W. (1956). The effects of monetary reward and punishment on auditory perception. *The Journal of Psychology*, *41*(1), 177–184. <https://doi.org/10.1080/00223980.1956.9916214>
- Stanners R. F., Coulter M., Sweet A. W., & Murphy P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion*, *3*(4), 319–340. <https://doi.org/10.1007/BF00994048>
- Strand J. F., Brown V. A., & Barbour D. L. (2020). Talking points: A modulating circle increases listening effort without improving speech recognition in young adults. *Psychonomic Bulletin & Review*, *27*(3), 536–543. <https://doi.org/10.3758/s13423-020-01713-y>
- Strand J. F., Brown V. A., Merchant M. B., Brown H. E., & Smith J. (2018). Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *Journal of Speech Language and Hearing Research*, *61*(6), 1463–1486. [https://doi.org/10.1044/2018\\_Jslhr-H-17-0257](https://doi.org/10.1044/2018_Jslhr-H-17-0257)
- Strauss D. J., & Francis A. L. (2017). Toward a taxonomic model of attention in effortful listening. *Cognitive Affective & Behavioral Neuroscience*, *17*(4), 809–825. <https://doi.org/10.3758/s13415-017-0513-0>
- Suurmond R., van Rhee H., & Hak T. (2017). Introduction, comparison and validation of meta-essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, *8*(4), 537–553. <https://doi.org/10.1002/jrsm.1260>
- Thompson S. G., & Higgins J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*(11), 1559–1573. <https://doi.org/10.1002/sim.1187>
- Tune G. S. (1966). Errors of commission as a function of age and temperament in a type of vigilance task. *Quarterly Journal of Experimental Psychology*, *18*(4), 358–361. <https://doi.org/10.1080/14640746608400054>
- Tune S., Wostmann M., & Obleser J. (2018). Probing the limits of alpha power lateralisation as a neural marker of selective attention in middle-aged and older listeners. *European Journal of Neuroscience*, *48*(7), 2537–2550. <https://doi.org/10.1111/ejn.13862>
- van Doorn R. R. A., & Lang J. W. B. (2010). Performance differences explained by the neuroticism facets withdrawal and volatility, variations in task demand, and effort allocation. *Journal of Research in Personality*, *44*(4), 446–452. <https://doi.org/10.1016/j.jrp.2010.05.004>
- Wallace J. F., & Newman J. P. (1997). Neuroticism and the attentional mediation of dysregulatory psychopathology. *Cognitive Therapy and Research*, *21*(2), 135–156. <https://doi.org/10.1023/A:1021828628571>
- Wallace J. F., & Newman J. P. (1998). Neuroticism and the facilitation of the automatic orienting of attention. *Personality and Individual Differences*, *24*(2), 253–266. [https://doi.org/10.1016/S0191-8869\(97\)00153-0](https://doi.org/10.1016/S0191-8869(97)00153-0)
- Weis T., Puschmann S., Brechmann A., & Thiel C. M. (2013). Positive and negative reinforcement activate human auditory cortex. *Frontiers in Human Neuroscience*, *7*, 842–842. <https://doi.org/10.3389/fnhum.2013.00842>
- Winn M. B., Wendt D., Koelewijn T., & Kuchinsky S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An Introduction for those Who want to Get started. *Trends in Hearing*, *22*, 2331216518800869. <https://doi.org/10.1177/2331216518800869>
- Wisniewski M. G., Thompson E. R., & Iyer N. (2017). Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly

- difficult. *Psychophysiology*, 54(12), 1916–1928. <https://doi.org/10.1111/psyp.12968>
- Zekveld A. A., Koelewijn T., & Kramer S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, 22, 1–25. <https://doi.org/10.1177/2331216518777174>
- Zekveld A. A., van Scheepen J. A. M., Versfeld N. J., Veerman E. C. I., & Kramer S. E. (2019). Please try harder! The influence of hearing status and evaluative feedback during listening on the pupil dilation response, saliva-cortisol and saliva alpha-amylase levels. *Hearing Research*, 381, 107768. <https://doi.org/10.1016/j.heares.2019.07.005>
- Zhang M. (2017). *Listening Effort Allocation: Stimulus-Driven, Goal-Driven, or Both?* [Doctoral dissertation]. University of Pittsburgh, Pittsburgh, USA. Retrieved from <http://d-scholarship.pitt.edu/32459/>
- Zhang M., Siegle G. J., McNeil M. R., Pratt S. R., & Palmer C. (2019). The role of reward and task demand in value-based strategic allocation of auditory comprehension effort. *Hearing Research*, 381, 1–15. <https://doi.org/10.1016/j.heares.2019.107775>