



The meta-analyses of deliberate practice underestimate the effect size because they neglect the core characteristic of individualization—an analysis and empirical evidence

Tobias Debatin¹ · Manuel D. S. Hopp² · Wilma Vialle³ · Albert Ziegler²

Accepted: 17 September 2021
© The Author(s) 2021

Abstract

Influential meta-analyses have concluded that only a small to medium proportion of variance in performance can be explained by deliberate practice. We argue that the authors have neglected the most important characteristic of deliberate practice: individualization of practice. Many of the analyzed effect sizes derived from measures that did not assess individualized practice and, therefore, should not have been included in meta-analyses of deliberate practice. We present empirical evidence which suggests that the level of individualization and quality of practice (indicated by didactic educational capital) substantially influences the predictive strength of practice measures. In our study of 178 chess players, we found that at a high level of individualization and quality of practice, the effect size of structured practice was more than three times higher than that found at the average level. Our theoretical analysis, along with empirical results, support the claim that the explanatory power of deliberate practice has been considerably underestimated in the meta-analyses. The question of how important deliberate practice is for individual differences in performance remains an open question.

Keyword Deliberate practice · Talent development · Skill acquisition · Expertise

Introduction

Despite a body height of only 1.81 m, Stefan Holm became the world champion (indoor) and the Olympic champion in high jump. The documentary film “Im Körper der Topathleten [In the body of top athletes]” (Yano & Miyano, 2008) shows how these great achievements were made possible. Holm compensated for his natural body height limitation (for a professional high jumper) with an extensive individualized training schedule from the time he was a young child. His story is a

prime example of what is possible when best learning practices are applied over extended periods of time.

Ericsson et al. (1993) coined the term ‘deliberate practice’ to describe this type of optimal learning practice. They made two important claims about its significance for expertise development, one related to intra-individual skill development and the other to inter-individual differences. With regard to intra-individual skill development, Ericsson et al. (1993) claimed that “high levels of deliberate practice are necessary to attain expert level performance” (p. 392). It is not a sufficient condition, however, because deliberate practice activities can also be associated with failure, through overtraining, for example. Successful attempts to continuously push individual limits would require problem solving, finding effective learning strategies and practice tasks, designing structured practice in an optimal learning setting, and setting appropriate learning goals. Thus, there is no guarantee that even well-designed attempts to push one’s limits upwards will be successful, let alone other forms of practice in a domain, such as play, mere experience,

✉ Tobias Debatin
tobias.debatin@ur.de

¹ Department of Educational Sciences, University of Regensburg, Regensburg, Germany

² Department of Educational Psychology, Friedrich-Alexander University Erlangen-Nuremberg, Nuremberg, Germany

³ Faculty of Social Sciences, University of Wollongong, Wollongong, Australia

or mindless drill. Such activities would likely lead to plateaus in skill acquisition. With regard to inter-individual differences, Ericsson et al. (1993) made the bold claim that “individual differences in ultimate performance can largely be accounted for by differential amounts of past and current levels of practice” (p. 392).

Meta-Analyses of Deliberate Practice

Macnamara et al. (2014) and Hambrick et al. (2014) posited that although deliberate practice is necessary in the acquisition of expertise, it might be a far less powerful factor than originally proposed. In the most comprehensive meta-analysis of deliberate practice research to date, Macnamara et al. (2014) reported that only 19% of variance in performance, averaged over different domains and even after correcting for estimates of measurement error, was accounted for by deliberate practice. Clearly, 19% seems incompatible with the assumption of performance differences being largely accounted for by deliberate practice. While the comprehensive meta-analysis of Macnamara et al. (2014) was concerned with many domains, the earlier published meta-analysis of Hambrick et al. (2014) only analyzed the domains of chess and music. They performed a reanalysis of the existing studies of chess and music expertise, corrected for measurement error, and found—when assuming a reliability of deliberate practice estimates of 0.80—34% and 30% of variance in chess and music performance, respectively, that could be explained by accumulated hours of deliberate practice. Additionally, they reported a high variability in accumulated hours of practice among the chess players at certain levels of chess expertise. Based on these findings, the authors argued that the deliberate practice framework is not sufficient to explain performance differences between individuals and proposed greater focus on other explanatory constructs, such as general cognitive ability.

However, as will be outlined below, it is highly questionable whether the authors of the meta-analyses did justice to the concept of deliberate practice. As there is currently no broadly accepted definition of deliberate practice and consequently a lot of confusion about the term, in the next section we will start by highlighting the core elements of the concept according to Ericsson and colleagues. In our opinion—despite the lack of a single accepted definition—it is not problematic to define the characteristics that are necessary for practice to be considered as deliberate practice. Most importantly, to our knowledge, there has not been any doubt that individualization of practice is necessary (e.g., informative feedback and diagnosis of errors). It is true that there are controversies and even contradictions regarding the need of a coach or a teacher, but we will save this (secondary) discussion for later.

The Core Characteristics of Deliberate Practice and Distinguishing it from Other Types of Practice

Ericsson et al. (1993) pointed out that deliberate practice is not simply any deliberate learning activity, but rather those “activities that have been found most effective in improving performance” (Ericsson et al. (1993), p. 367). In the section on “Characteristics of deliberate practice” (p. 367) they describe in detail what is crucial in order to determine that practice is deliberate practice.

The first important characteristic, described by Ericsson et al. (1993), was that “deliberate practice is a highly structured activity, the explicit goal of which is to improve performance” (p. 368). In line with Ericsson and Harwell (2019), we will use the term ‘structured practice’ to describe any structured practice that is aimed at improvement. This term is intended to denote an incomplete operationalization, which does not measure deliberate practice. Additionally, we will adopt the term ‘naive practice’ to describe any unstructured practice or activity that is not aimed at improvement and, therefore, is even lacking the first characteristic of deliberate practice (Ericsson & Harwell, 2019). Ericsson and colleagues emphasized that naive practice will not lead to substantial improvements of skills, if any improvement at all (e.g., Ericsson, 2013, 2014; Ericsson et al., 1993).

A second core characteristic is individualized task construction and informative feedback. The authors make it clear that the use of adequate learning strategies and methods is essential for improvement and to avoid being stuck at a particular level. But the critical question is how you get optimal feedback and how you find the most suitable learning strategies for you. Ericsson et al. (1993) reiterate that ideally you need teachers and coaches to achieve that difficult goal. Nevertheless, even with a teacher there is still the question of whether the teacher is capable of recognizing strengths and weaknesses and can provide precise feedback as well as personalized strategies to overcome those weaknesses. Consequently, it seems clear that in order to regard activities as deliberate-practice activities, the quality of the learning activities has to be high, irrespective of whether it is with a (good) teacher or through careful monitoring and individualization of your own learning process (Nandagopal & Ericsson, 2012).

Later in the text, Ericsson et al. (1993) define deliberate practice as “a highly structured activity, the explicit goal of which is to improve performance. Specific tasks are invented to overcome weaknesses, and performance is carefully monitored to provide cues for ways to improve it further. We claim that deliberate practice requires effort and is not inherently enjoyable” (p. 368). Our observations suggest that the authors’ directives—in relation to tasks aimed at overcoming weaknesses and monitoring for further

improvement—are often overlooked and, consequently, deliberate practice is predominantly associated with being highly structured, aimed at improving skills, and not being inherently enjoyable.

While a high degree of structure and the explicit goal of improvement are certainly core characteristics of deliberate practice, they are not sufficient because individualization of practice is also assumed to be necessary in deliberate practice. Thus, theoretical considerations and empirical research seem to indicate a need to distinguish among at least three types of activities: naive practice, structured practice (with the aim of improvement), and individualized practice. The last of these also needs to be aimed at improvement, as well as being carefully and – most importantly – competently designed and regulated to determine the most suitable learning pathway for a specific individual. Ericsson does not explicitly suggest the term ‘individualized practice’, but in more recent publications Ericsson uses ‘purposeful practice’ to refer to individualized practice without a coach and ‘deliberate practice’ to refer to individualized practice with a coach (Ericsson & Harwell, 2019).

Reexamining the Meta-Analyses

Having these necessary core characteristics in mind is important as the subjective main inclusion criterion in the meta-analyses of Macnamara et al. (2014) and Hambrick et al. (2014) is that a study has to contain an “activity interpretable as deliberate practice”. As we will outline in this section, the operationalization of deliberate practice in the selected studies is often far from what is intended by this term (this is not a critique of these studies as they often did not claim to measure deliberate practice; for a comprehensive reanalysis of the meta-analysis see Ericsson and Harwell (2019)). For example, Loyens et al., (2007, p. 585) simply asked students in each of eight courses to report the “mean number of hours spent on self-study per week”. The eight correlations of this variable with grades in the final exam of each course ranged from $r=0.02$ to $r=0.24$ and were included as eight effect sizes in the meta-analysis of Macnamara et al. (2014). As another example, the study of Howard (2012) with the largest sample size in the domain of chess, and which contributed the smallest effect size of all chess studies ($r=0.33$), simply asked in a survey, “How many hours per week on average have you studied chess since taking up the game seriously?”. This estimate was multiplied with the number of years of serious practice to get an estimate of total study hours and correlated with the recent chess rating. Accordingly, the authors made it clear that this is no measure of deliberate practice: “no claims about deliberate practice are made here” (Howard, 2012, p. 360).

As could be expected from these examples, the reanalysis of Ericsson and Harwell (2019) found a much higher uncorrected percentage of explained variance than did Macnamara et al. (2014) (29% vs. 14%). In this reanalysis, Ericsson and Harwell only included effects of practice measures which could be defined as deliberate practice or at least as purposeful practice. Correcting for realistic estimates of reliability, the explained amount of variance can easily result in an estimated amount of more than 50% of explained variance. The corrections of Ericsson led to an explained amount of variance of 61%. At this point we note that we do not endorse another reanalysis of Miller et al. (2020) which found a meta-analytic estimate that was only slightly higher than that of Macnamara et al. (2014). We agree with their theoretical critique but they still included effect sizes from simple study hours.

Overall, it seems that the theoretical foundations of the deliberate practice framework leave substantial room for improvement in the prediction of performance differences other than the correction of measurement error alone. We would argue that the meta-analyses largely confirm the starting assumption of Ericsson et al. (1993) that a large amount of practice does not necessarily lead to improvement. As suggested in the meta-analyses, other factors likely play a role in the development of expertise. Nevertheless, it is important to consider factors within the deliberate practice framework before attributing all of the unexplained variance to other factors.

Current Study

The main aim of our study was to quantify how important the core characteristics of deliberate practice, especially the characteristic of individualization, are for predicting chess skill development over the course of one year.

For this purpose, we assessed the construct of didactic educational capital (didactic EC), which is defined as “[...] the assembled know-how involved in the design and improvement of educational and learning processes” (Ziegler & Baker, 2013, p. 29) to which an individual has access. It is closely related to the characteristics of deliberate practice and even more so to the more specific characteristic of individualization, since the availability of good feedback opportunities is an important component of didactic EC. Therefore, the higher the didactic EC of an individual is, the closer their practice activities with a clear aim of improvement (structured practice) should be to deliberate practice. Assessing the didactic EC of a person thus makes it possible to evaluate if the predictive strength of structured practice increases the closer it

comes to deliberate practice, as indicated by higher levels of didactic EC.

Based on the theoretical assumptions of the introduction, we arrived at the following hypotheses:

H1a: Structured practice is positively related to chess skill development.

Due to well-known findings about non-linear learning curves in skill acquisition, which also apply to the domain of chess (Howard, 2014), we additionally assumed as a secondary hypothesis:

H1b: At lower skill levels the same amount of structured practice leads to more improvement in chess skill than in higher skill levels.

H2: The predictive strength of structured practice increases substantially the higher the didactic EC of an individual is.

Method

Procedure and Sample

The data set consisted of German-speaking tournament chess players who were recruited via the chess platform, ChessBase (<https://de.chessbase.com>). There they completed an online questionnaire which assessed demographics, their chess ratings, when they played their most recent rated game, as well as information about their chess practice and everyday activities. For more details about the measurements, see the Measures section. No ethics approval was required for this kind of study by our institutional review board. The sample is the same as in Debatin et al. (2015) but there is no conceptual overlap as this study focused on parts of the online questionnaire not used in the present study, namely the everyday activities of the players. Due to our hypotheses and quality demands on the data set, we narrowed down the initial 219 tournament chess players to a sample of 178 players through the following four steps:

- (1) to obtain a measure of the change in chess skill during the past year, we excluded all tournament chess players who did not report a chess rating at one year previously or reported not having played a rated chess game during the last year (excluded 23 players).
- (2) we excluded players who gave unrealistic high estimates of their practice time, i.e., more than 168 h per week which corresponds to 24 h per day (excluded three players).
- (3) to ensure high data quality we excluded players who had obvious language problems (e.g., that entered phrases or words which made no sense written in the space for estimated hours and other fields; excluded four players).
- (4) The remaining 189 tournament chess players showed some missing values, i.e., 1.15% of all observations were missing with a maximum of 4.76% in the age variable, which translated into 178 complete cases (excluded eleven players, i.e., 5.82% which most likely leads to unbiased results (Graham, 2009)).

The data of the remaining 178 tournament chess players (171 men, 7 women) were used in the following analyses.

Measures

Chess Skill Chess skill was assessed via Elo rating (Elo, 1987). The Elo rating is an international, objective chess skill measure, which predicts tournament success very well (Hambrick et al., 2014).

Participants reported their most recent Elo rating and their rating at one year previously. The latter was termed “Elo T1”. The Elo score’s high correlation of .91 with tournament success (Hambrick et al., 2014) suggests that reliability should be acceptable even as a self-report measure.

Structured Practice Participants were asked to estimate the hours they spent weekly on playing chess currently and one year previously. Participants were then asked to report how many of these hours (currently and one year previously) were “serious practice with the aim of improvement”. The mean of the latter two assessments (i.e., mean of current and previous year serious practice with the aim of improvement) was termed “Structured Practice”.

The reliability of self-reported cumulative life-time practice is typically found to be around 0.75 (Côté et al., 2005; Hambrick et al., 2014). Bilalić et al. (2007) found correlations of 0.98 and 0.99 between training diary entries and retrospective estimates (6 months) of the amount of practice in young chess players. Thus, the reliability of the self-reported estimates of the previous year in our study should exceed the lifetime estimates significantly.

Didactic Educational Capital The degree of individualization and quality of chess practice was assessed with the additional question: “I receive high quality, individualized training for developing my chess skills (either) in chess club, from chess partners or otherwise.” The question is part of an unpublished short scale for assessing Educational Capital (Ziegler & Baker, 2013) in the domain of chess. A five-point Likert-type scale ranging from 1 = “not at all true”

to 5 = “totally true” was used. This variable was termed “Didactic EC”.

Research in the field of education has consistently shown the validity of evaluation procedures (Stehle et al., 2012; Wachtel & Wachtel, 1998). Stehle et al. (2012), for example, found that even a one-item question (as used in our study) concerning the overall quality of the course was a good—although not ideal—predictor ($r=0.50$) of the outcome of a practical examination following the course.

Plan of Analysis

For our regression analysis, the difference between the most recent Elo rating and the Elo rating one year previously was used as the dependent variable and was termed “Elo Change”. As recommended by Cohen et al. (2003), we used these change scores while including the Elo rating at one year previously as a predictor variable. Concerning the effects of other predictors, this procedure leads to identical results (regarding unstandardized coefficients, test statistics and p values) as predicting the recent Elo rating and including the Elo rating at one year previously as a predictor variable (Werts & Linn, 1970). Using the Elo Change variable just simplifies the interpretation of the results.

Due to the high prevalence the occurrence of outliers (see Table 2) and to avoid subjective judgments about how to deal with them, we rejected the use of ordinary least squares (OLS) linear regression estimation since it is highly sensitive to outliers (Anderson & Schumacker, 2003; Wilcox, 2017). Robust regression tackles this problem—simply put—by decreasing the weight of highly influential outliers. There are different methods of robust regression, we chose MM-type robust regression, developed by Yohai (1987) as implemented in robustbase (Maechler et al., 2020), which uses a bi-square redescending score function and returns highly robust and highly efficient estimates. MM-estimation outperforms ordinary least squares (OLS) estimation in the presence of outliers (Anderson & Schumacker, 2003; Wilcox, 2017). For assessing the explained variance (adj.) R^2 of Elo Change, we used the consistency corrected robust coefficient of determination by Renaud and Victoria-Feser

(2010) and for assessing change in R^2 we used the robust Wald test (Ghosh et al., 2016).

We used four increasingly complex robust multiple linear regression models. Our first step contained Age, Elo T1 and Structured Practice as independent variables to address Hypothesis 1a. Our second step addressed Hypothesis 1b that at lower skill levels the same amount of structured practice leads to more improvement in chess skill than in higher skill levels by including the interaction effect Elo T1 \times Structured Practice. Our third and fourth step included Didactic EC and its interaction with Structured Practice which led to our final model: a multiple regression of Elo Change on Age, Elo T1, Structured Practice, Elo T1 \times Structured Practice, Didactic EC, and Didactic EC \times Structured Practice. With our fourth step we tested Hypothesis 2.

All variables were first standardized and then the multiplicative terms were calculated from the z -scores to get a proper standardized solution as recommended in Cohen et al. (2003). Standardizing (as centering) of the original variables prevents nonessential multicollinearity between the multiplicative terms and the original variables (Cohen et al., 2003).

All our analyses were conducted with R v4.0.2 for general analysis (R Core Team, 2020) and utilized the following packages: psych v2.0.9 for descriptives and correlations (Revelle, 2020), rstatix v0.6.0 for outlier detection (Kassambara, 2020), robustbase v0.93–6 for the robust regression (Maechler et al., 2020), and ggplot2 v3.3.2 for visualizations (Wickham, 2016).

Results

Outlier and Correlation Analysis

The descriptive statistics can be found in Table 1. The outlier analysis of Elo Change, Structured Practice, Didactic EC, Elo T1 and Age found outliers (i.e., above quartile three + $1.5 \times$ inter-quartile range or below quartile one – $1.5 \times$ inter-quartile range) in Elo Change, Structured Practice and Elo T1 (see Table 2). Elo Change revealed 28

Table 1 Descriptives

	<i>M</i>	<i>SD</i>	Median	MAD	Min	Max	Skew	Kurtosis
Elo Change	6.32	49.10	0.00	17.79	-100	380	3.55	21.06
Structured Practice	3.33	6.00	1.50	2.22	0	40	4.02	18.54
Didactic EC	2.54	1.27	2.00	1.48	1	5	0.41	-0.86
Elo T1	2015.92	241.53	2042.50	211.27	1100	2670	-0.72	1.83
Age	41.15	13.22	42.50	14.08	13	78	0.09	-0.52

$N=178$, MAD stands for Median Absolute Deviation. EC stands for Educational Capital

Table 2 Outliers

Variable	Number of outliers*	Number of extreme outliers**
Elo Change	28	11
Structured Practice	13	10
Didactic EC	0	0
Elo T1	8	0
Age	0	0

* Values above quartile three + $1.5 \times \text{IQR}$ or below quartile one — $1.5 \times \text{IQR}$, ** Values above quartile three + $3 \times \text{IQR}$ or below quartile one — $3 \times \text{IQR}$. EC stands for Educational Capital

outliers, of which 11 were classified as extreme (i.e., above quartile three + $3 \times$ inter-quartile range or below quartile one — $3 \times$ inter-quartile range). Structured Practice uncovered 13 outliers, of which 10 were classified as extreme, and Elo T1 showed eight non-extreme outliers. Age and Didactic EC did not show any outliers. A visualization of the outliers can be found in Fig. 1, where boxplots of the standardized variables Elo Change, Structured Practice and Didactic EC are presented.

To identify correlations between the variables, we chose Spearman's rank-order correlations since this method is relatively robust to outliers (Croux & Dehon, 2010; de Winter et al., 2016). All estimates can be found in Table 3. Change in Elo shows small correlations with Didactic Educational Capital ($r_S(176) = 0.16$, $p = 0.047$) and with Structured Practice ($r_S(176) = 0.19$, $p = 0.014$), thus providing initial evidence for Hypothesis 1a. We found a negative correlation between Elo Change and Age ($r_S(176) = -0.32$, $p < 0.001$) which indicates that younger chess players—on

the average—have a higher change in their Elo scores over a one-year period than older chess players.

Hierarchical Linear Regression Results

We tested four increasingly complex multiple linear regression models with the robust MM-estimation (see Table 4). In Model 1, we included the control variables Elo T1 and age as well as structured practice. We then added the interaction Elo T1 \times Structured Practice (model 2) and Didactic EC (model 3) consecutively. Our final model 4 included the interaction of Didactic EC and Structured Practice.

In model 1, Elo Change was positively related to Structured Practice ($\beta = 0.09$, $t = 2.53$, $p = 0.012$). Structured Practice added $\Delta R^2 = 0.024$ ($p = 0.011$) to the model including only Age and Elo T1. Thus, we accept Hypothesis 1a.

Model 2 included the interaction effect of Elo T1 and Structured Practice which improved model fit ($\Delta R^2 = 0.12$, $p < 0.001$) and showed a negative beta weight ($\beta = -0.13$, $p < 0.001$). This model indicates that chess players with low Elo T1 score—on average—increased their Elo score more than chess players with high Elo T1 score. Thus, we accept Hypothesis 1b.

Model 3 included Didactic EC, our measurement of individualization and quality of practice, which neither increased model fit ($\Delta R^2 < 0.01$, $p = 0.26$) nor showed a significant beta weight ($\beta = 0.05$, $p = 0.26$).

The final model 4 increased the explained variance to $R^2 = 0.43$ (adj. $R^2 = 0.41$, $\Delta R^2 = 0.15$, $p = 0.005$) and the added interaction term of Didactic EC and Structured Practice showed a positive beta weight ($\beta = 0.37$, $p = 0.005$). Figure 2 shows this interaction. In simple terms, average chess players in our sample (concerning Elo T1) with a high level of individualization and quality of practice (i.e., +1 SD

Fig. 1 Boxplots of standardized variables. Note. EC stands for Educational Capital

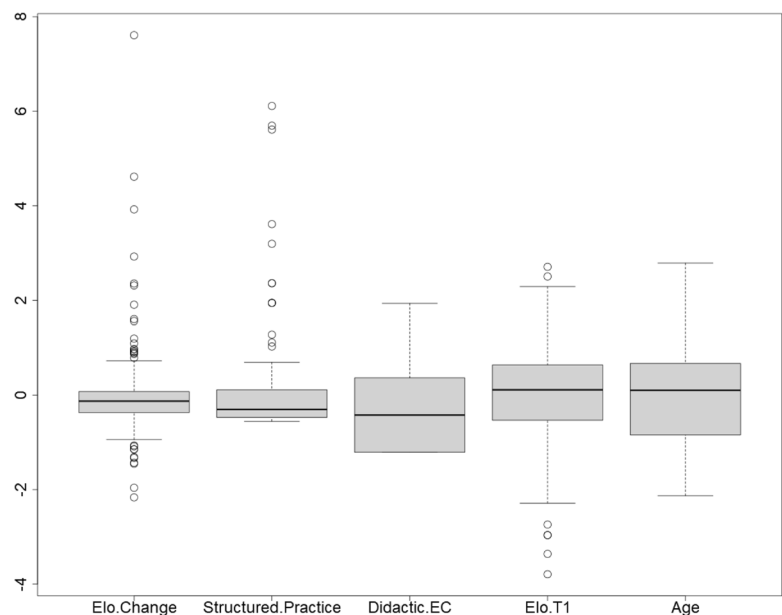


Table 3 Spearman's rank-order correlations (N=178)

	1	2	3	4	5
Elo Change	–				
Structured Practice	0.19* [0.04, 0.33]	–			
Didactic EC	0.16* [0.00, 0.30]	0.27*** [0.12, 0.42]	–		
Elo T1	-0.11 [-0.26, 0.04]	0.03 [-0.13, 0.19]	-0.02 [-0.18, 0.13]	–	
Age	-0.32*** [-0.46, -0.16]	-0.06 [-0.21, 0.09]	-0.17* [-0.31, -0.02]	-0.05 [-0.20, 0.11]	–

* Indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$. Bootstrapped 95% CI are displayed in brackets. EC stands for Educational Capital

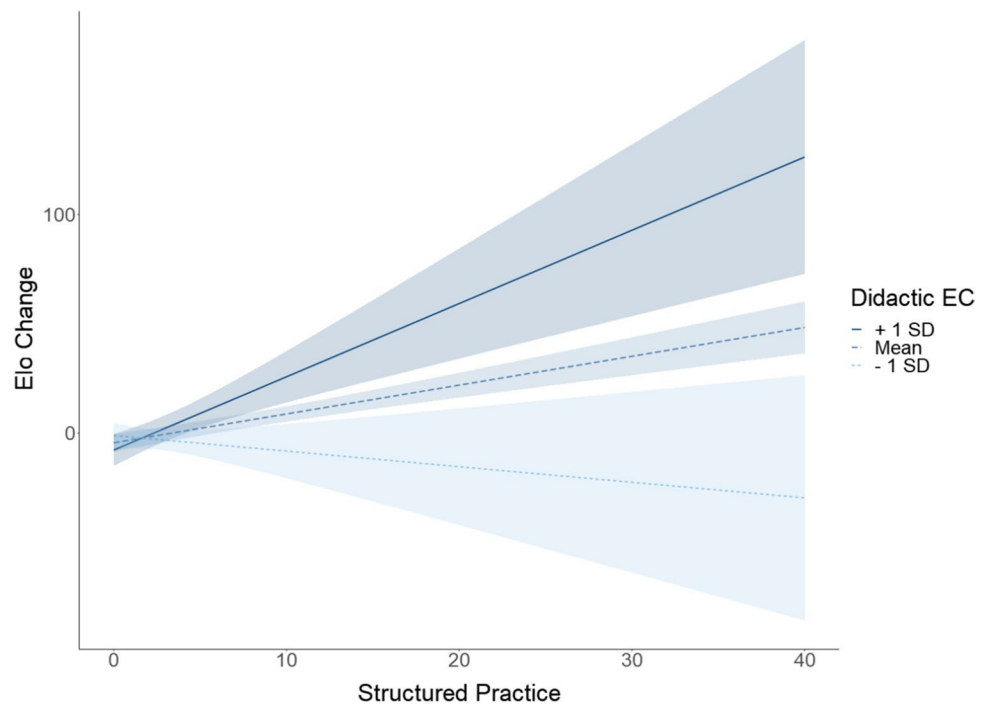
Table 4 Results of robust hierarchical multiple regression of Elo Change

Predictor	beta	beta 95% CI [LL, UL]	t value	Fit	Difference
(Intercept)	-0.12**	[-0.20, -0.05]	-3.31		
Age	-0.16**	[-0.27, -0.06]	-3.21		
Elo T1	-0.05	[-0.13, 0.04]	-1.10		
Structured Practice	0.09*	[0.02, 0.17]	2.53		
				$R^2 = 0.15$	
				Adj. $R^2 = 0.14$	
(Intercept)	-0.10**	[-0.18, -0.03]	-2.88		
Age	-0.15**	[-0.26, -0.06]	-3.14		
Elo T1	0.05	[-0.13, 0.03]	-1.27		
Structured Practice	0.24***	[0.17, 0.31]	7.06		
Elo T1 × Structured Practice	-0.13***	[-0.19, -0.08]	-4.80		
				$R^2 = 0.27$	$\Delta R^2 = 0.12^{***}$
				Adj. $R^2 = 0.26$	
(Intercept)	-0.11**	[-0.18, -0.03]	-2.83		
Age	-0.16**	[-0.25, -0.06]	-3.14		
Elo T1	-0.05	[-0.13, 0.03]	-1.21		
Structured Practice	0.23***	[0.15, 0.31]	5.58		
Elo T1 × Structured Practice	-0.14***	[-0.20, -0.08]	-4.45		
Didactic EC	0.05	[-0.03, 0.13]	1.13		
				$R^2 = 0.27$	$\Delta R^2 < 0.01$
				Adj. $R^2 = 0.25$	
(Intercept)	-0.11**	[-0.18, -0.04]	-3.00		
Age	-0.16**	[-0.26, -0.06]	-3.06		
Elo T1	-0.05	[-0.12, 0.02]	-1.37		
Structured Practice	0.16***	[0.12, 0.20]	7.63		
Elo T1 × Structured Practice	-0.41***	[-0.63, -0.19]	-3.64		
Didactic EC	0.07	[-0.01, 0.15]	1.63		
Didactic EC × Structured Practice	0.37**	[0.11, 0.62]	2.83		
				$R^2 = 0.43$	$\Delta R^2 = 0.15^{**}$
				Adj. $R^2 = 0.41$	

Beta indicates the standardized regression weights. LL and UL indicate the lower and upper limits of the 95% confidence interval, respectively. EC stands for Educational Capital.

*Indicates $p < 0.05$. ** indicates $p < 0.01$. *** indicates $p < 0.001$

Fig. 2 Interaction plot of centered variables. Note. EC stands for Educational Capital



Didactic EC) improve more than three times faster through structured practice than chess players with average levels of individualization and quality of practice (i.e., average Didactic EC). Thus Hypothesis 2 was confirmed as the predictive strength of structured practice substantially increases, the higher the didactic EC of an individual is.

Discussion

On a theoretical level and integrating the recent suggestions of Ericsson and Harwell (2019), we argued the need to distinguish between different types of practice: naive practice, structured practice and individualized practice (an umbrella term we use for Ericsson's terms purposeful and deliberate practice, both of which he characterizes as involving individualization, the difference between the two terms being whether there is the direct support of a teacher). In line with Ericsson and Harwell (2019) we define naive practice as an unstructured practice activity without the clear aim of improvement and structured practice as a structured practice activity that explicitly focuses on skill improvement. We further argued that only individualized practice, which means structured practice that is additionally characterized by high quality, such as individual feedback and specific, competently designed learning tasks (with or without a teacher), can be a strong predictor of individual differences in skill acquisition and consequently in skill levels.

The empirical results of our study clearly confirm this proposal. At high levels of individualization and quality of practice (indicated by didactic EC), the effect size of structured practice was more than three times higher than that of an average level of individualization and quality of practice.

We also showed that it is no problem to define the necessary core characteristics of deliberate practice, independent of the (secondary) debate of whether a teacher is necessary. Most importantly, to our knowledge, there has not been any doubt that individualization of practice is necessary (e.g., informative feedback and diagnosis of errors). However, 'necessary' is not sufficient and it is true Ericsson made contradictory claims concerning the need of a teacher for deliberate practice. Nonetheless, in meta-analyses of deliberate practice (Hambrick et al., 2014; Macnamara et al., 2014), the practice activities in the analyzed studies often did not fulfill the necessary characteristics of deliberate practice, especially regarding individualization, and therefore it is safe to say that the meta-analyses clearly underestimate its effect size. Again, in line with Ericsson, we think it is more appropriate to say that the meta-analyses are analyzing structured practice.

These findings have important implications for the question of whether deliberate practice can explain a large portion of expertise differences, as discussed in Macnamara et al. (2014) and Hambrick et al. (2014). Our results indicate that the strength of prediction of practice activities is considerably weakened if the individualization and quality of practice is not assessed as in many studies included in

the meta-analyses. We do not think a teacher or coach is essential for the design of good practice activities; it seems possible for individuals to monitor and individualize the learning process quite well on their own (Nandagopal & Ericsson, 2012). However, we assume it is very likely that a good personal teacher facilitates the acquisition of expertise more readily. Therefore, we think Ericsson's most recent statements that deliberate practice must involve a teacher or coach should be adopted, despite previous contradictions.

Overall, we think it is certainly possible that individual differences in performance can largely be accounted for (interpreted as explaining more than half of the variance) by different amounts of accumulated deliberate practice hours. Nevertheless, we think there is considerable variance left which can be explained by other factors.

One of the most discussed candidates is (general) intelligence. There is indeed evidence that intelligence is positively related to skill acquisition, predominantly in novel and relatively complex tasks (e.g., Ackerman, 1988; Voelkle et al., 2006). However, there are findings which are important to highlight in this context. Several studies showed an overlap between intelligence and certain (teachable) metacognitive skills and that the unique effect of intelligence on learning was smaller than the unique effect of metacognitive skills (Veenman & Spaans, 2005; Veenman et al., 2004). Considering that metacognitive skills are basically part of the deliberate practice framework, effects of intelligence might not be situated completely outside of the framework of deliberate practice. This is also illustrated by a small study which showed an effect of intelligence on learning poker in a discovery learning condition, but not in a guided discovery learning condition (DeDonno, 2016). It seems reasonable to assume that intelligence helps "to build your own individualized practice" when not receiving instruction, probably via increased metacognitive skills. However, this does not exclude the influence of biological variables: for example, recent intelligence research suggests that higher levels of energy metabolism in the brain facilitate learning (Debatin, 2019, 2020). Another finding to consider from intelligence research is that effects on learning seem to be weaker or not even present when considering long-term learning outside the laboratory (for an overview of growth curves studies see Debatin et al., 2019). We think the reason is that the development of expertise should be seen from a dynamic systems perspective, in which intelligence is only one of many variables that interact over time. Overall, the role of intelligence in skill acquisition is certainly not simple as already described in the theory of Ackermann (1988).

Another important factor in skill acquisition seems to be the availability of external resources to which an individual has access (e.g., Ziegler & Baker, 2013). Ericsson et al. (1993) explicitly addressed the role of resources such

as the financial investment of parents and the encouragement provided by them. It seems fair to summarize the view of Ericsson et al. by saying that the effect of resources on skill development is (predominantly) mediated by deliberate practice hours. Indeed, there is empirical evidence for a relationship between external resources in the domain of chess (most strongly for the social resources) and the amount of time playing chess (Debatin et al., 2015). However, we think there might also be rather strong direct effects of different kinds of resources, meaning they are not necessarily mediated by deliberate practice. For example, when your close personal contacts like to talk about chess, it might provide positive memory effects outside of conscious training. A particular effective method for improving several resources simultaneously might be mentoring, especially when a deep personal connection develops between the mentee and the mentor (Stoeger et al., 2019).

In concluding this paper, we want to emphasize that in the future more focus should be on the assessment of individualization and quality of (deliberate) practice activities to get a clearer picture of the importance of deliberate practice for skill development. The limitations of our study point to more specific directions for future research. An obvious limitation is the broad and subjective estimation of the variable of individualization and quality of practice. However, though weak as a measurement, it concomitantly strengthens our point. If it was possible to find clear indications that individualization and quality matters with an obviously sub-optimal measure, it seems plausible to assume that the explained variance would be even higher if the quality of practice could have been assessed in a more comprehensive way. Additionally, the reliability of the change in Elo rating is unclear since we do not know the number of games that influenced the Elo development during the previous year. Further, we cannot exclude the possibility that players who had improved over the previous year were biased in retrospectively reporting a more positive view of the quality of their training; nevertheless, this seems unlikely to be a major concern because we only found a weak association between didactic EC and Elo change. Instead, the interaction of didactic EC and structured practice showed a rather strong effect. The next obvious research step should be to replicate our results in a predictive longitudinal design with a more elaborated measurement instrument for the degree of individualization and quality of practice activities.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The datasets generated during and/or analysed during the current study as well as the code are available from the corresponding author on reasonable request.

Declarations

Conflicts of Interest/Competing Interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics Approval No ethics approval was required for this kind of study by our institutional review board.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*(3), 288–318. <https://doi.org/10.1037//0096-3445.117.3.288>
- Anderson, C., & Schumacker, R. E. (2003). A comparison of five robust regression methods with ordinary least squares regression: Relative efficiency, bias, and test of the null hypothesis. *Understanding Statistics*, *2*(2), 79–103. https://doi.org/10.1207/S15328031US0202_01
- Bilalić, M., McLeod, P., & Gobet, F. (2007). Does chess need intelligence? - A study with young chess players. *Intelligence*, *35*(5), 457–470. <https://doi.org/10.1016/j.intell.2006.09.005>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum Associates Inc.
- Côté, J., Ericsson, K. A., & Law, M. P. (2005). Tracing the development of athletes using retrospective interview methods: A proposed interview and validation procedure for reported information. *Journal of Applied Sport Psychology*, *17*(1), 1–19. <https://doi.org/10.1080/10413200590907531>
- Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, *19*(4), 497–515. <https://doi.org/10.1007/s10260-010-0142-z>
- de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, *21*(3), 273–290. <https://doi.org/10.1037/met0000079>
- Debatin, T. (2019). A revised mental energy hypothesis of the g factor in light of recent neuroscience. *Review of General Psychology*, *23*(2), 201–210. <https://doi.org/10.1177/1089268019832846>
- Debatin, T. (2020). Neuroenergetics and “General Intelligence”: A systems biology perspective. *Journal of Intelligence*, *8*(3), 31. <https://doi.org/10.3390/jintelligence8030031>
- Debatin, T., Harder, B., & Ziegler, A. (2019). Does fluid intelligence facilitate the learning of English as a foreign language?—A longitudinal latent growth curve analysis. *Learning and Individual Differences*, *70*, 121–129. <https://doi.org/10.1016/j.lindif.2019.01.009>
- Debatin, T., Hopp, M., Vialle, W., & Ziegler, A. (2015). Why experts can do what they do: The effects of exogenous resources on the Domain Impact Level of Activities (DILA). *Psychological Test and Assessment Modeling*, *57*(1), 94–110.
- DeDonno, M. A. (2016). The influence of IQ on pure discovery and guided discovery learning of a complex real-world task. *Learning and Individual Differences*, *49*, 11–16. <https://doi.org/10.1016/j.lindif.2016.05.023>
- Elo, A. E. (1987). The rating of chess players, past and present. In *New York: Arco*.
- Ericsson, K. A. (2013). Training history, deliberate practice and elite sports performance: An analysis in response to Tucker and Collins review-what makes champions? *British Journal of Sports Medicine*, *47*(9), 533–535. <https://doi.org/10.1136/bjports-2012-091767>
- Ericsson, K. A. (2014). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticisms. *Intelligence*, *45*(1), 81–103. <https://doi.org/10.1016/j.intell.2013.12.001>
- Ericsson, K. A., & Harwell, K. W. (2019). Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research. *Frontiers in Psychology*, *10*(OCT), 1–19. <https://doi.org/10.3389/fpsyg.2019.02396>
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. <https://doi.org/10.1037/0033-295X.100.3.363>
- Ghosh, A., Mandal, A., Martín, N., & Pardo, L. (2016). Influence analysis of robust Wald-type tests. *Journal of Multivariate Analysis*, *147*, 102–126. <https://doi.org/10.1016/j.jmva.2016.01.004>
- Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, *60*(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence*, *45*(1), 34–45. <https://doi.org/10.1016/j.intell.2013.04.001>
- Howard, R. W. (2012). Longitudinal effects of different types of practice on the development of chess expertise. *Applied Cognitive Psychology*, *26*(3), 359–369. <https://doi.org/10.1002/acp.1834>
- Howard, R. W. (2014). Learning curves in highly skilled chess players: A test of the generality of the power law of practice. *Acta Psychologica*, *151*, 16–23. <https://doi.org/10.1016/j.actpsy.2014.05.013>
- Kassambara, A. (2020). *rstatix: Pipe-friendly framework for basic statistical tests* (0.6.0).
- Loyens, S. M. M., Rikers, R. M. J. P., & Schmidt, H. G. (2007). The impact of students' conceptions of constructivist assumptions on academic achievement and drop-out. *Studies in Higher Education*, *32*(5), 581–602. <https://doi.org/10.1080/03075070701573765>
- Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate Practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science*, *25*(8), 1608–1618. <https://doi.org/10.1177/0956797614535810>
- Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibián-Barrera, M., Verbeke, T., Koller, M., Conceicao, E., & Di, Palma, M. (2020). *robustbase: basic robust statistics*.
- Miller, S. D., Chow, D., Wampold, B. E., Hubble, M. A., Del Re, A. C., Maeschalck, C., & Bargmann, S. (2020). To be or not to be (an expert)? Revisiting the role of deliberate practice in improving performance. *High Ability Studies*, *31*(1), 5–15. <https://doi.org/10.1080/13598139.2018.1519410>

- Nandagopal, K., & Ericsson, K. A. (2012). An expert performance approach to the study of individual differences in self-regulated learning activities in upper-level college students. *Learning and Individual Differences, 22*(5), 597–609. <https://doi.org/10.1016/j.lindif.2011.11.018>
- R Core Team. (2020). R: A language and environment for statistical computing reference index. In *R Foundation for Statistical Computing*. <https://doi.org/10.1109/ICIN.2011.6081064>
- Renaud, O., & Victoria-Feser, M.-P. (2010). A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference, 140*(7), 1852–1862. <https://doi.org/10.1016/j.jspi.2010.01.008>
- Revelle, W. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University.
- Stehle, S., Spinath, B., & Kadmon, M. (2012). Measuring teaching effectiveness: Correspondence between students' evaluations of teaching and different measures of student learning. *Research in Higher Education, 53*(8), 888–904. <https://doi.org/10.1007/s11162-012-9260-9>
- Stoeger, H., Debatin, T., Heilemann, M., & Ziegler, A. (2019). Online mentoring for talented girls in STEM: The role of relationship quality and changes in learning environments in explaining mentoring success. *New Directions for Child and Adolescent Development, 2019*(168), 75–99. <https://doi.org/10.1002/cad.20320>
- Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences, 15*(2), 159–176. <https://doi.org/10.1016/j.lindif.2004.12.001>
- Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14*(1), 89–109. <https://doi.org/10.1016/j.learninstruc.2003.10.004>
- Voelkle, M. C., Wittmann, W. W., & Ackerman, P. L. (2006). Abilities and skill acquisition: A latent growth curve approach. *Learning and Individual Differences, 16*(4), 303–319. <https://doi.org/10.1016/j.lindif.2006.01.001>
- Wachtel, H. K., & Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education, 23*(2), 191. <https://doi.org/10.1080/0260293980230207>
- Werts, C. E., & Linn, R. L. (1970). A general linear model for studying growth. *Psychological Bulletin, 73*(1), 17–22.
- Wickham, H. (2016). *ggplot2 elegant graphics for data analysis*.
- Wilcox, R. (2017). *Modern statistics for the social and behavioral sciences*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315154480>
- Yano, S., & Miyano, H. (directors). (2008) *Im Körper der Topathleten [In the body of top athletes]* [Film]. Arte.
- Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics, 15*(2), 642–656. <https://doi.org/10.1214/aos/1176350366>
- Ziegler, A., & Baker, J. (2013). Talent development as adaptation: The role of educational and learning capital. In S. Phillipson, H. Stoeger, & A. Ziegler (Eds.), *Exceptionality in East Asia: Explorations in the Actiotope Model of Giftedness* (pp. 18–39). Routledge.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.