E-JOURNAL Bewegung und Training

www.ejournal-but.de

ORIGINAL ARTICLE

DOES FREQUENT AUGMENTED FEEDBACK REALLY DEGRADE LEARNING? A META-ANALYSIS

FRANZ MARSCHALL[1], ANDREAS BUND[2], & JOSEF WIEMEYER[2]
[1]*UNIVERSITÄT DES SAARLANDES*, [2]*TECHNISCHE UNIVERSITÄT DARMSTADT*

Die Häufigkeit und Verteilung (Frequenz) ergänzender Rückinformation hat einen bedeutenden Einfluss auf das motorische Lernen. Empirische Befunde legen die Vermutung nahe, dass eine reduzierte Informationsfrequenz die Aneignung von Bewegungen negativ, das Lernen im Sinne eines relativ überdauernden Behaltens dagegen positiv beeinflusst. Die vorliegende Metaanalyse prüft diesen „Umkehreffekt" über die Auswertung von insgesamt 40 Studien. Auf der Grundlage des Zufall-Effekt-Modells werden verschiedene meta-analytische Techniken (vote-counting, globale Analyse und Moderatoranalyse) angewendet. Der vermutete Unkehreffekt lässt sich nur als schwacher globaler Effekt nachweisen. Die Moderatoranalyse zeigt, dass er allein in Verbindung mit Laboraufgaben, mit der Aufgabenstellung Parameteroptimierung und bei sehr hohen Übungsfrequenzen auftritt. Insgesamt hat eine reduzierte Frequenz von Rückinformation keine gegenüber vollständiger Rückinformation unterschiedliche Wirkung. Unter Anwendungsgesichtspunkten ist damit ein Argument geliefert, auf die ständige Bereitstellung von Rückinformation zu verzichten. Forschungsmethodisch sollten an Stelle der resultatsbezogenen abhängigen Variablen trial-to-trial-Analysen verwendet werden, um den Prozess der Informationsverarbeitung besser untersuchen zu können.

Schlüsselwörter: Ergebnisrückmeldung, Häufigkeit der Ergebnisrückmeldung, Guidance-Hypothese, motorisches Lernen

The frequency of augmented feedback is generally considered to have an essential influence on motor learning. Empirical evidence suggests that reducing information frequency degrades acquisition, but enhances learning. The purpose of this meta-analysis is to examine this reversal effect by analyzing and coding a sample of 40 studies. We applied the meta-analytic techniques of vote-counting, global, and moderator analysis based on the random effects model. Our results indicate that the reversal effect may be an artifact resulting from the use of specific research methods. The reversal effect was not confirmed as a global phenomenon. Prerequisites for the emergence of the reversal effect were the use of particular laboratory tasks, parameter learning, and extensive practice. We discuss the implications of these findings for practice and research methodology. The missing negative effect of reduced augmented feedback should be interpreted as "exoneration": It is not necessary to give feedback after every practice trial in acquisition. Instead of ex post facto analysis, a trial-to-trial research method should be applied as an appropriate method for investigating information processing strategies dependent on particular augmented feedback conditions.

Keywords: knowledge of results, frequency of augmented feedback, guidance hypothesis, motor learning

Introduction

Besides practice, augmented information in terms of knowledge of results (KR) or knowledge of performance (KP) is considered an essential variable influencing motor learning (for a summary, see Blischke, Marschall, Müller, & Daugs, 1999; Magill, 2001; Swinnen, 1996). The timing, precision, frequency, and type of knowledge of results are often manipulated in order to arrange training and education in an efficient way. The appropriate frequency and distribution of augmented information, in order to enhance performance and learning, seems to be rather controversial. Many recommendations exist (for a review, see Magill, 2001; Wulf, 1992), especially those based on evidence that reveals a reversal effect (Vickers, 1994; Wiemeyer, 1998). That is to say that reduced frequency of augmented feedback degrades acquisition performance but enhances learning, whereas full feedback generally has the opposite effect. However, several studies do not confirm this reversal effect and instead indicate no degradation in acquisition performance under reduced frequency conditions (Weeks & Sherwood, 1994) and/or no detriment to learning effects with full feedback (Lai & Shea, 1999; Sparrow & Summers, 1992; Winstein & Schmidt, 1990, Exp. 1). The importance of a lack of consistency for the efficacy of reduced frequency KR schedules goes beyond empirical shortcomings but has implications for contemporary theory regarding the role of information feedback for learning. Specifically, the guidance hypothesis (see Schmidt, 1991) and the feedback usefulness hypothesis (see Wulf, 1994) could not be confirmed (Dunham & Mueller, 1993; Marschall, Blischke, & Müller, 1997; Marschall, Müller, & Blischke, 1997). Even within single studies there are mixed results that both support and favour the reversal effect (e.g., Wulf & Schmidt, 1989).

The lack of clarity regarding the influence of scheduling of information feedback for performance and learning provided the impetus for the present analysis. The purpose of the proposed

Correspondence Address:
Franz Marschall (f.marschall@mx.uni-saarland.de)
Sportwissenschaftliches Institut
Universität des Saarlandes
Postfach 151150, 66041 Saarbrücken
Germany

E-Journal *Bewegung und Training*
Offizielles Organ der Sektionen Biomechanik, Sportmotorik und Trainingswissenschaft in der Deutschen Vereinigung für Sportwissenschaft
ISSN 1612-5770

meta-analysis is a quantitative summary of studies addressing the impact of different frequencies of augmented feedback for performance and learning. We examined the assumption that reduced feedback frequency (as compared to full [100%] frequency) degrades acquisition performance, but enhances learning, particularly in late retention. In the following sections we first introduce the different variants of reduced feedback scheduling and the effect on motor learning. Second, in order to clarify our methodology, we then describe acquisition, coding, and judgement of the studies and the determination of the moderator variables. Finally, we present the results comprising vote-counting, global analysis, and moderator analysis.

Different Frequencies of Augmented Feedback: Phenomenon and Evidence

Literature unequivocally defines frequency of augmented feedback: "Absolute frequency refers to the absolute number of times in a learning sequence that KR is provided to the learner. Relative frequency is defined as the absolute frequency of KR divided by the total number of trials given, and it expresses the proportion of trials for which KR was provided" (Salmoni, Schmidt, & Walter, 1984, p. 362). However, this concept is applied to quite different informational conditions. For example, one could distinguish between feedback frequency and feedback distribution, and then consider the reversal effect in each case.

Relative frequency of augmented feedback is the ratio of feedback trials and total trials and is usually expressed as a percentage (e.g., Winstein & Schmidt, 1990). This ratio can be applied as a fixed ratio, a random schedule, or a fading schedule (i.e., more frequent feedback in the early acquisition and increasingly less frequent feedback in the later acquisition phase; Marschall, 1992, p. 41). One might consider the type of feedback manipulation as a form of analogue to a well documented practice schedule manipulation (massed versus distributed practice; Schmidt, 1982, p. 482-484.). With respect to augmented feedback, a distributed presentation (permanent change of trials with and without feedback) is compared to a massed or blocked presentation (blocks of trials with and without feedback).

Alternatively, there exist particular manipulations of KR that do not necessarily reduce or eliminate information content. For example, when applying bandwidth feedback, augmented feedback is only delivered if performance is beyond a special range of tolerance (e.g., Goodwin & Meeuwsen, 1995). Depending on the range of tolerance, different frequencies and distributions of augmented feedback result. However, this is only a superficial information reduction because no feedback implies that the performance was within the respective boundaries of tolerance and can therefore serve as reinforcement. Summary KR or KP is a second particular form of reduced feedback. In this case, augmented feedback is presented only after a certain number of practice trials have been completed. The number of practice trials without feedback depends on the number of summarized trials (see, for example, Guadagnoli, Dornier, & Tandy, 1996). Summary feedback also delivers information for every acquisition trial (Sidaway, Moore, & Schoenfelder-Zohdi, 1991). Identic constraints are given, when KR is delayed over trials (Anderson, Magill, Sekiya & Ryan, 2005).

While both frequency and distribution schedules may vary, an important characteristic that appears to lead to the reversal effect described here is the learner's exposure to strings of trials with and without KR. The study of Goodwin and Meeuwsen (1995) using different bandwidth (BW) variants may serve as an example to illustrate this point. Participants learned golf putting under four different KR conditions (for which the acquisition phase included 10 blocks of 10 trials): (1) KR for every trial (BW 0%), (2) KR only for trials when target distance was exceeded by more than 10% (BW 10%), (3) shrinking BW (range of tolerance decreasing gradually from 20% to 5% during acquisition), and (4) expanding BW (range of tolerance increasing gradually from 0% to 20%). The group with most frequent KR (BW 0%) performed best in acquisition, whereas the group with the lowest amount of KR (BW 10%) performed best in immediate and late retention.These results confirm earlier evidence that frequent augmented feedback enhances acquisition and degrades learning and vice versa (Goodwin & Meeuwsen, 1995, p. 102).

Method of Meta-Analysis

Localisation, Coding, and Judgement of the Studing

Meta-Analysis provides a resolution to decide on conflicting evidence of various outcomes in research findings (R. Rosenthal & DiMatteo, 2001) and is called "the state-of-the-art procedure for the quantitative synthesis of research findings across studies" (Hagger, 2006).

The lack of clarity regarding the influence of scheduling of information feedback for performance and learning provided the impetus for the present meta-analysis. It is based on a total of 40 studies and was executed 2001. We retrieved the studies by way of a search of two relevant literature databases (*Spolit* and *Psychlit*), Psychological Abstracts, and two significant scientific journals (*Journal of Motor Behavior* and *Research Quarterly for Exercise and Sports*). Furthermore, we searched according to the snowball principle (M. C. Rosenthal, 1994). From 132 sources selected, 84 references were considered relevant for the meta-analysis. Finally, we were able to include 40 studies in the meta-analysis. Criteria for inclusion were existing statistical values (mean and standard deviation for experimental and control group, *t* values, and *F* values; R. Rosenthal, 1994). When the respective values were missing for more than 50% of the single results, the study was excluded from the meta-analysis. However, the vote-counting method could be used for the excluded studies (see Table 2).

The quality of study (Rustenbach, 2003) was judged according to five criteria: The validity of the dependent measure, internal validity, situation validity, population validity, and statistical validity (Schlicht, 1994). Each of the five criteria was assigned a maximum possible score of 4 points (very weak = 1 point, weak = 2 points, good = 3 points, excellent = 4 points), and each was evaluated by two independent persons. The quality of study was then operationally defined as the mean of the five criteria scores. The interpersonal reliability of the judgement was satisfactory (r = .84). Therefore, quality of study was included as a moderator variable. The results of the quality judgement show that the studies are generally of good quality (M = 2.63 points; SD = 0.46): 14 are "good/very good" (3-4 points), and 4 are "weak" (2 points).

The validity of the dependent measure depends on the quality of assessment and the relation of dependent measure on the one hand, and the task and feedback information on the other. Validity is reduced if measures of consistency are assessed although the task explicitly requires exact reproduction. Many studies measure several variables without testing whether these measures are appropriate and independent. No adjustment of significance level takes place in case of non-orthogonality (see, for example, Bortz, Lienert, & Boehnke, 1990, p. 48); validity is also reduced if no multivariate statistical data analysis is performed, something that is actually required due to error measures not being independent (Schmidt, 1982). High internal validity is achieved if an experimental design is applied and control variables are assessed. Usually gender is controlled and taken into account as covariate, but only three studies assess cognitive or motor variables as control variables. Situation and population validity is reduced in most of the studies because general hypotheses are tested using restricted samples (usually students, in most cases taking part in athletic activities). In not one of the studies are sample sizes based on appropriate calculations utilizing significance level, power and effect size (Thomas, Lochbaum, Landers, & He, 1997). In some studies, sample size varies for no apparent reason (Gable, Shea, & Wright, 1991; Guay, Salmoni, & McIlwain, 1992). Statistical validity is reduced in most of the studies. Only 12 studies report complete results and only two studies report effect sizes (*ES*). In the rest of the studies means and standard deviations or F and t values are not reported. These studies were rated "very weak" (1 point) or "weak" (2 points). However, it seems to be much more critical that no statistical values are reported with non-significant results and that non-significant results are not discussed (see, for discussion, R. Rosenthal & DiMatteo, 2001, p. 63). Furthermore, the possible relation between significant main effects and interactions is rarely addressed (e.g., Goodwin & Meeuwsen, 1995). All these problems lead to reduced scores for statistical validity.

Moderator Variables

The purpose of this meta-analysis was the estimation of reliable population *ES* on the basis of homogeneous data. If these preconditions cannot be confirmed, it is necessary to take moderators into account in order to clarify residual variance (Schlicht, 1994, p. 53). Moderators can therefore be considered systematic sources of variance (Rustenbach, 2003, p. 186). Drinkmann (1990, p. 91) proposes that moderator variables show a relation to the studied effect, and that they should not only be determined based on practical considerations. The examination of moderator variables also adds to theory development and increases the importance of the meta-analysis (R. Rosenthal & DiMatteo, 2001). Based on this proposal we determined the following moderators.

Quality of study. We divided the studies based on the total score of the five criteria discussed above. We scored each criterion on a 4-point scale (Schlicht, 1994). Good studies earned a score above 2.5 and weak studies a score below 2.5 points.

Number of acquisition trials. Variability in the number of trials may have an influence on motor learning and also on the emergence of the reversal effect. Therefore, four categories were included in meta-analysis: 1 to 30 trials, 31 to 60 trials, 61 to 90 trials, and more than 90 trials. These categories were chosen to achieve similar frequencies.

Reduction of feedback. The most important purpose of this meta-analysis was to compare a 100% feedback treatment to reduced feedback treatments. However, there was a considerable range of feedback reduction conditions. Most often, studies reduced feedback from 33% to 66% of the respective total acquisition trials. We compared studies with 0 to 33% feedback reduction with studies using 34 to 66% reduction and did not locate any study with more than 66% feedback reduction. We expected that greater feedback reduction would lead to a reversal effect because this condition may result in more pronounced reduction of acquisition performance.

Feedback distribution. According to Sidaway et al. (1991) we distinguish practice conditions where no feedback is given for single trials, i.e. a real reduction of feedback (fixed ratio, fading, random, or self-controlled feedback), from conditions including only a reduction of feedback presentation (i.e., bandwidth, summary feedback). When providing bandwidth feedback there is information pertaining to every trial, either as direct feedback in the case of exceeding the tolerance limits or as indirect feedback of a nearly correct movement in the case of missing feedback. When providing summary feedback, information is also delivered for every trial but with different temporal delays. Because of the significance of "blank trials" (Swinnen, 1996, p. 53) there may be different effects of real information reduction when compared to reduction of only the information presentation (bandwidth, summary feedback).

Content of feedback. Although different frequency of augmented feedback is the independent measure in the studies examined, a closer look at this variable shows that there are substantial differences concerning the informational content associated with particular feedback conditions. Feedback varies from simple KR information (e.g., "Your movement time was 3.7 seconds") to discrepancy information between actual and desired outcome "You moved one second too fast!" or complete KP and corrective feedback "Your movement time was 3.7 seconds. You have to slow down 1 second on the next trial!" (Kernodle & Carlton, 1992). It is possible that frequency effects may be more pronounced if simple KR is the only information source.

Type of feedback. In research on feedback there is a distinction between KR (knowledge of results) and KP (knowledge of performance; see, for example, Schmidt, 1982, p. 426). Because these different feedback types convey different information, they may moderate the reversed frequency effect in a different way.

Type of task. We distinguished different types of task based on the respective relative timing. According to the Generalized Motor Program (GMP) theory (Schmidt, 1982), tasks in which a defined feature within the framework of a mastered movement structure is approximated to a specific parameter value to achieve the optimum effect were categorized as parameter learning. In contrast, tasks in which the specific spatio-temporal progression of the movement is to be learned were considered as program learning. Many studies confirming a reversal effect require parameter learning tasks, so this parameter learning may improve with reduced feedback frequency.

Task context. We further delineated studies on the basis of the task context. We specifically distinguished between laboratory tasks (usually small artificial movements like positioning a lever) and sport tasks (usually sport-specific or sport-related movements like golf putting). Swinnen (1996, p. 40) has suggested that reversed frequency effects may be more evident in cases in which internal and external feedback sources are minimized. This is more likely the case with laboratory tasks, hence the expectation of the reversed frequency effect in laboratory tasks only.

Data Analysis

Study effect size (*ES*) of the d family were included in the meta-analysis (R. Rosenthal, 1994; R. Rosenthal & DiMatteo, 2001). We calculated *ES* based on the data of the primary studies (see Table 1). Frequently more than one *ES* were calculated because more than two experimental groups were examined or more than one dependent variable was analyzed. Therefore, the number of *ES* was greater than the number of studies (acquisition: 72 *ES* from 25 studies; early retention: 64 *ES* from 22 studies; late retention: 80 *ES* from 25 studies). We summarized the *ES* by calculating the study *ES* means (Hagger, 2006, p. 113). The summarized data were processed by the meta-analysis program of Schwarzer (1997; Meta-analysis 5.3).

Table 1

Coding of Studies

Study (Year)	*N*	Feedback reduction (%)	Measure	Mean study effect size		
				Acquisition	Early retention	Late retention
1. Behrman et al. (1992)	16	50	Performance	-0.94	0.94	0.94
2. Blischke et al. (1993)	40	50	Performance	-0.69	-0.49	-0.18
3. Broker, Gregor, & Schmidt (1993)	20	5,5	Performance	-0.21		-0.02
4. Butler & Fischman (1996)	22	25	Performance	-0.38/-0.38	0.69/0.29	-0.68/-0.08
5. Butler & Fischman (1999)	28	22; 28	Performance		0.67	
6. Butler, Reeve, & Fischman (1996)	20	–	Consistency	0.41	0.19	0.19
7. Carnahan et al. (1996)	48	20	Performance/Consistence	-1.37/0.22	0.69/0.01	
8. Cauraugh, Chen, & Radlo (1993)	24	65	Performance/Consistence	0.01/-0.09	-0.12/-0.42	
9. Gable, Shea, & Wright (1991)	24	12,5; 6,25	Performance			0.72
10. Goodwin & Meeuwsen (1995)	60	–	Performance/Consistence	-0.45/0.45		
11. Greuter (1996)	41	50; 25	Performance	0.17	0.18	-0.09
12. Guay et al. (1992, Exp. 1)	20	20; 10; 6,6	Performance	-0.82		
13. Guay et al. (1992, Exp. 2)	40	20; 10	Performance	-0.55		
14. Herbert & Landin (1994)	24	0	Performance	-1.05		-1.61
15. Hillebrecht (1994, Exp. 1)	40	50	Performance/Consistence	-0.33/0.03	-0.29/-0.61	
16. Hillebrecht (1994, Exp. 2)	60	50	Performance/Consistence	0.08/-0.05	-0.07/-0.35	0.03/-0.32
17. Hillebrecht & Schuster (1994)	27	50	Performance/Consistence	-0.10/-0.12	-0.13/-0.10	-0.29/-0.16
18. Jarus (1995)	30	33	Performance		0.77	

Table 1

Coding of Studies (Continued)

Study (Year)	*N*	Feedback reduction (%)	Measure	Mean study effect size		
				Acquisition	Early retention	Late retention
19. Lai , Shea, Wulf, & Wright (2000)	40	50	Performance/Consistence	0.71/0.96		3.11/4.67
20. Lee & Maraj (1994)	20	–	Performance		0.76	
21. Marschall (1992)	59	50; 25	Performance	-0.22	-0.03	
22. McCullagh & Little (1990)	30	33	Performance/Consistence	-0.07/-0.58	0.41/-0.18	0.00/-0.45
23. Nicholson & Schmidt (1991)	58	50	Performance			0.45
24. Schlicher (1993)	40	50	Performance	0.34	-0.02	0.30
25. Schmidt et al. (1990, Exp. 1)	30	20; 10; 6,6	Performance	-0.64		0.64
26. Schmidt et al. (1990, Exp. 2)	39	20	Performance			0.72
27. Schmidt et al. (1989)	36	20; 10; 6,6	Performance	-0.58		
28. Sidaway et al. (1991)	24	50	Performance	-0.73	-0.73	-0.73
29. Smith, Taylor, & Withers (1997)	16	–	Consistency			0.78
30. Sparrow & Summers (1992)	26	0; 10; 20; 33	Performance		0.70	
31. Weeks & Sherwood (1994)	30	20	Consistency	0.64		
32. Wiemeyer (1998, Exp. 2)	20	33; 73	Performance	-0.06		-0.83
33. Winstein et al. (1994)	20	33	Performance		-1.20	0.82
34. Winstein & Schmidt (1990)	58	50	Performance			0.67
35. Wulf (1992)	48	66	Performance		-0.49	
36. Wulf (1994, Exp. 9)	34	66	Performance		0.03	0.42
37. Wulf & Schmidt (1989, Exp. 1)	26	66	Performance/Consistence			0.93/0.60
38. Wulf & Schmidt (1989, Exp. 2)	30	66	Performance/Consistence		0.63	0.79/0.70
39. Wulf et al. (1993, Exp. 1)	38	63	Performance			0.23
40. Wulf et al. (1993, Exp. 2)	38	63	Performance	0.68		

Meta-analyses can be performed based on two different statistical models. The fixed-effect model (Hedges, 1994) is based on the assumption of a fixed common population *ES*. This model should be able to account for all the sources of variance. The assumption of a fixed common *ES* may be rarely true because many studies may be subject not only to systematic variance induced by the treatment but also random variance that can not be accounted for by the model (Field, 2003). The random effects model (Raudenbush, 1994) takes this into account and is based on the assumption of random distribution of population *ES*. It was therefore considered the more appropriate model (Hagger, 2006). According to the random effects model we calculated the mean *ES* (Δ), the respective significance (*Z*) and homogeneity (*Q*). We considered *ES* of 0.2 as weak, 0.5 as moderate and 0.8 as strong (Cohen, 1992). However, tests for homogeneity should be considered carefully because the power is relatively small, particularly with a small number of primary studies and with small sample sizes of the separate studies. Furthermore we calculated the variance due to sample error. A value of 100% indicates that the observed *ES* variance is completely due to sample error. Additionally, we calculated Orwin's Fail Safe N for *ES* of 0.2, 0.5 and 0.8 to address the file drawer problem (Hagger, 2006, p. 107; Rustenbach, 2003, p. 248).

Results

Vote-Counting

The results of published studies usually show distortion because evidence that does not support the hypothesis and that is referred to as "non-significant results" is often not interpreted and documented incompletely or even not at all (Bliesener, 1999). The study of Wulf and Schmidt (1989) provides an example of this problem. In this study comparing the frequency of 100% KR and 67% KR, a total of 23 single results are reported. Only six results are significant. From these significant results only one result can be considered as confirmation of the reversal effect. However, in the subsequent discussion, the authors generalize this single result as providing support for program learning. This example indicates that the reversal effect and the guidance hypothesis may be inappropriately over-interpreted when based on single results.

These incompletely documented studies are usually not included in a meta-analysis. This is why studies that do not support their respective hypotheses are under-represented as compared to studies supporting their hypotheses. Therefore meta-analyses are inherently biased: The actual population effect is overestimated based on a non-representative sample.

In order to test this bias we used a vote-counting method (Rustenbach, 2003, p. 200). We classified the results of the studies either as confirming or not confirming the hypothesis according to the reversal effect, i.e., better acquisition performance of the 100% group and better learning of the reduced frequency group. Non-significant results and significant results in the opposite direction were classified as not confirming.

We included 40 studies with 72 single results for acquisition and retention performance with existing statistical values (mean and standard deviation for experimental and control group, *t* values, *F* values). Based on the above mentioned criteria, 44 studies with 94 single results were not included because the respective values were missing for more than 50% of the single results. The large number of total studies and single results are due to the fact that multiple considerations in acquisition, early and late retention were conducted. Of course, there is a change of category possible, e.g., if a study can only be included for acquisition, but not for early and late retention (e.g., Guay et al., 1992), or if a study confirmed the hypothesis in the acquisition but not in the retention phase (e.g., Broker, Gregor, & Schmidt, 1993) or vice versa.

The results presented in Table 2 confirm the sample bias. Of the 25 studies included for acquisition, 14 studies confirmed the hypothesis whereas 11 studies did not. From the 43 studies that could not be included because of missing statistical values, 39 studies did not confirm the hypothesis, whereas only 4 studies did. The results are similar for early and late retention. The results of Chi-square tests confirm this bias, acquisition: $\chi^2 = 17.71$, $p < .001$; early retention: $\chi^2 = 11.76$, $p < .001$; late retention: $\chi^2 = 9.09$, $p < .001$.

From a statistical point of view we find a Bayes error because the probability for inclusion in the meta-analysis under the precondition of confirming results is not equal to the inclusion probability under the precondition of non-confirming results. In order to avoid these sample errors future review procedures should take more care to include complete documentation of non-significant results. As a consequence of this bias we calculated the Fail Safe N for effect sizes of 0.2, 0.5 and 0.8. The Fail Safe N gives the number of studies with an effect size (*ES*) of zero that is necessary in order to reduce the effect size below 0.2, 0.5 or 0.8 (Ashworth, Osburn, Callendar, & Boyle, 1992; Rustenbach, 2003, p. 248). By comparing the Fail Safe N and the results of our vote counting we can test whether the required number of zero-effect studies actually exists or not.

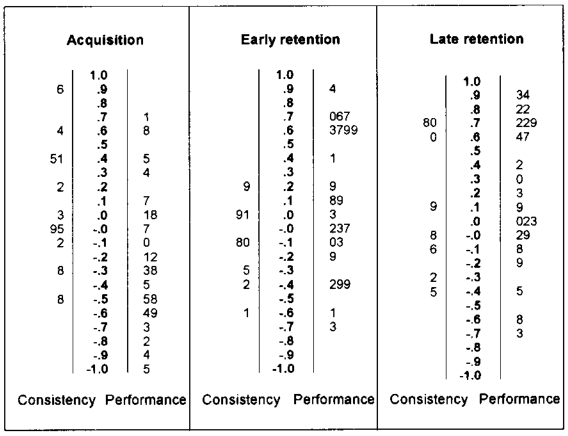

Figure 1. Stem-and-leaf plots of acquisition, early retention, and late retention community.

Global Analysis

To get a first impression of the effects of full versus reduced feedback frequency we performed a global analysis of all effect sizes. We only distinguished between acquisition and early and late retention. In the acquisition phase subjects practice the skill under different feedback conditions; in early and late retention learning is tested in the absence of augmented feedback. Early retention tests are performed after a pause lasting only a few minutes, whereas late retention requires an interval of at least 24 hours. Table 3 shows the results of this analysis. For acquisition we find a significant negative population effect for performance measures ($\Delta = -0.21$, $Z = -2.19$, $p < .05$). That is, 100% feedback frequency groups show better acquisition results as compared to groups with reduced feedback frequency. However, this effect is not homogenous ($Q = 35.45$, $p < .05$) and is reduced to below 0.2 (low significance for practice) if only two zero-effect studies exist. The results of our vote-counting confirm that many more studies do indeed exist (see Table 2). The population effect for the consistency measures is not significant ($\Delta = 0.15$, $Z = 1.18$, $p > .05$). Sampling error variance is approximately 65%.

Table 2
Vote-Counting of Relevant Studies

	Acquisition		Early retention		Late retention		Σ
	+	–	+	–	+	–	
Included studies	14	11	11	11	15	10	72
Not included studies	4	39	2	26	4	19	94
Σ	18	50	13	37	19	29	

+ confirming the hypothesis; – not confirming the hypothesis

Table 3
Results of the Global Analysis

Phase	Measure	*k*	Random effects model				
			Δ	*Z*	*Q*	Variance of sample error	Fail Safe N (0.8/0.5/0.2)
Acquisition	Performance	22	-0.21	-2.19*	35.45*	63.77%	0/0/2
	Consistency	11	0.15	1.18	15.34	66.14%	0/0/0
Early retention	Performance	20	0.15	1.48	32.58*	62.37%	0/0/0
	Consistency	8	-0.22	1.74*	3.96	100.00%	0/0/1
Late retention	Performance	21	0.22	2.08*	34.91*	53.92%	0/0/3
	Consistency	8	0.08	0.44	10.31	63.05%	0/0/0

* $p < .05$; ** $p < .01$; *** $p < .001$

In early retention the population effect for the performance measures is not significant ($\Delta = 0.15$, $Z = 1.48$, $p > .05$) with a tendency for better performance of the groups with reduced feedback frequency. For consistency we find a significantly negative effect ($\Delta = -0.22$, $Z = 1.74$, $p < .05$) indicating more consistent learning of the 100% groups. This effect size is homogenous ($Q = 3.96$, $p > .05$) but Fail Safe N is not acceptable because only one zero-effect study reduces the effect size below 0.2.

For the late retention phase, population *ES* of performance is significantly positive ($\Delta = 0.22$, $Z = 2.08$, $p < .05$) and heterogeneous ($Q = 34.91$, $p < .05$). Fail Safe N is again not acceptable with three zero-effect studies reducing the effect size below 0.2. For consistency in late retention we do not find a significant *ES* ($\Delta = 0.08$, $Z = 0.44$, $p > .05$). The reported results are illustrated by stem and leaf-plots (Figure 1). Outliers were excluded. Figure 1 shows that for acquisition most study *ES* are negative, whereas for early and late retention there is a tendency towards positive *ES*. However, in every phase there are also studies with poor or zero effects.

To summarize the results of the global analysis, we found heterogeneous results for the performance measures of all phases (acquisition, early, and late retention). For this measure a moderator analysis is suggested including the moderators that we discussed above in order to examine their contribution to *ES* variance. Possible sources of variance include methodological aspects of the study (quality, number of acquisition trials), information properties (reduction, distribution, content, and type of feedback information) and task properties (type and context). We performed separate meta-analyses for different stages of these variables while preserving the distinction of phases (acquisition, early, and late retention) and measures (performance, consistency).

Moderator Analysis

Quality of study. Based on the criteria proposed by Schlicht (1994) we judged studies with a mean total score below 2.5 (range of the scales: 1 to 4) as "poor" and studies with a score above 2.5 as "good". For good studies we find significant population *ES* for acquisition performance ($\Delta = -0.22$, $Z = -1.81$, $p < .05$) and early retention consistency ($\Delta = -0.31$, $Z = -2.04$, $p < .05$) indicating advantages for the 100% feedback groups. For poor studies we find a significant population *ES* for early retention performance ($\Delta = 0.39$, $Z = 2.34$, $p < .01$). All other *ES* are not significant. Considering the performance results without regard to significance we find that the reversal effect is confirmed by the poor studies (acquisition: $\Delta = -0.20$; early retention: $\Delta = 0.39$; late retention: $\Delta = 0.27$) rather than the good studies. With all *ES* the Fail Safe N values for 0.2 are not satisfactory, ranging from 0 to 5. Sample error variance ranges from 46% to 100%.

Number of acquisition trials. We assume that the number of acquisition trials has an important influence on the reversal effect because with increasing trials the influence of practice time and duration of feedback increases. Based on this assumption we devided the studies into four categories: 1 to 30 acquisition trials, 31 to 60 trials, 61 to 90 trials, and more than 90 trials. With one exception (early retention consistency: $\Delta = -0.38$, $Z = -2.31$, $p < .05$), we can only find significant *ES* with more than 30 acquisition trials. This holds true in particular for performance measures. For acquisition and early retention we find significant population *ES* with 31 to 60 trials ($\Delta = -0.39$, $Z = -1.35$, $p < .05$ and $\Delta = 0.48$, $Z = 2.36$, $p < .01$ respectively), 61 to 90 trials ($\Delta = -0.36$, $Z = -1.36$, $p < .05$ and $\Delta = -0.36$, $Z = -1.72$, $p < .05$), and more than 90 trials ($\Delta = -0.26$, $Z = -1.01$, $p < .05$ and $\Delta = 0.68$, $Z = 3.93$, $p < .001$). For late retention, significant results are confined to 61 to 90 and more than 90 trials ($\Delta = 0.31$, $Z = 1.53$, $p < .05$ and $\Delta = 0.55$, $Z = 2.30$, $p < .01$ respectively). With one exception (early retention performance with 61 to 90 acquisi-

Table 4

Hypotheses-Corresponding Significant Effect Sizes of the Moderator Analysis and Orwin's Fail Safe N for ES of 0.2

Moderator	Acquisition	Early retention	Late retention	Fail Safe N (0.2)		
	AC Δ	ER Δ	LR Δ	AC	ER	LR
Reduction of feedback, 0 – 33% (*k*=11)	-0.40	0.37		11	7	0
Acquisition trials 31-60 (*k*=3)	-0.39	0.48	0.19*	5	11	0
> 90 (*k*=5)	-0.26	0.68	0.55*	2	15	16
Schedule of feedback, reduced presentation (*k*=9)	-0.36	0.38	0.14*	8	5	0
Context of task, laboratory (*k*=16)	-0.25	0.27	0.38	4	6	4
Quality of study, poor studies (*k*=7)	-0.20	0.39	0.27*	1	5	2
Content of feedback, actual performance (*k*=5-8)	-0.25	0.24	0.34	2	2	6

tion trials), all the population *ES* confirm the reversal effect. However, some of the *ES* are not homogenous. In some cases, Fail Safe N reaches acceptable values (0 to 16 zero-effect studies for small effects, *ES* = 0.2). Sample error variance ranges from 28% to 100%.

Reduction of feedback. We compared studies with 0 to 33% feedback reduction to studies with 34 to 66% feedback reduction. Feedback reduction values less than 34% as compared to 100% feedback indicate reduced acquisition performance ($\Delta = -0.40$, $Z = -3.65$, $p < .01$), but enhanced early retention performance ($\Delta = 0.37$, $Z = 3.05$, $p < .01$). Fail Safe N values indicate high stability of this effect (acquisition: 11 zero-effect studies; early retention: 7), but this holds only for weak effect size (*ES* = 0.2). With respect to the 34% to 66% feedback reduction condition, early retention consistency is better under 100% feedback ($\Delta = -0.38$, $Z = -2.31$, $p < .05$), whereas late retention performance was enhanced under reduced feedback ($\Delta = 0.27$, $Z = 1.97$, $p < .05$). Based on these results the reversal effect may be confined to conditions that reduced feedback below 34%. Sample error variance ranges from 38% to 100%.

Feedback distribution. Following the suggestions of Sidaway et al. (1991), we can distinguish reduced feedback information (i.e., actually reduced feedback frequency) from reduced feedback presentation (i.e., bandwidth and summary feedback). We find significant ES for both reduced information and presentation cases. In acquisition and early retention reduced presentation frequency produces significant effects (acquisition performance: $\Delta = -0.36$, $Z = -2.54$, $p < .01$; acquisition consistency: $\Delta = 0.40$, $Z = 2.34$, $p < .01$; early retention performance: $\Delta = 0.38$, $Z = 1.69$, $p < .05$). In early and late retention we also find significant effects for reduced information (early retention consistency: $\Delta = -0.33$, $Z = -2.09$, $p < .05$; late retention performance: $\Delta = 0.30$, $Z = 2.89$, $p < .01$). The directions of these effects are unsystematically distributed and do not confirm a reversal effect. All effects are heterogeneous and Fail Safe N values range from 0 to 8 for a low *ES*. Sample error variance ranges from 46% to 100%.

Content of feedback. The studies apply different contents of feedback information: actual performance, discrepancy information, actual and desired performance, or transitional information. With one exception we could only analyze performance measures. In acquisition we found significant *ES* for actual and discrepancy information ($\Delta = -0.25$, $Z = -1.65$, $p < .05$ and $\Delta = -0.25$, $Z = -2.09$, $p < .05$). In late retention, however, the *ES* for actual performance and actual plus desired performance were significant ($\Delta = 0.34$, $Z = 1.77$, $p < .05$ and $\Delta = 0.40$, $Z = 1.87$, $p < .05$). All these *ES* were heterogeneous. The strong consistency effect for actual performance feedback in late retention ($\Delta = 0.63$, $Z = 2.30$, $p < .01$) is due to the two experiments of Schmidt, Young, Swinnen, & Shapiro (1989). In that case it is hard to interpret. Again, Fail Safe N values are low (0 to 6 zero-effect studies for low *ES*). Sample error variance ranges from 16% to 100%.

Type of feedback: Augmented feedback can be related to the movement outcome (KR) or procedural aspects of movement (KP). We find a significant population *ES* for KR in acquisition ($\Delta = -0.26$, $Z = -2.62$, $p < .01$) and for KP in late retention ($\Delta = 0.28$, $Z = 1.66$, $p < .05$). Fail Safe N for low *ES* ranges from 0 to 7 zero-effect studies. Sample error variance ranges from 39% to 76%.

Type of task. According to the concept of GMP (Schmidt, 1982), tasks can be differentiated in acquisition of a GMP (program learning) and acquisi tion of program parameters (parameter learning). As expected we find a reversal effect only with tasks requiring program learning or a combination of program and parameter learning. In acquisition, 100% feedback led to significantly better performance ($\Delta = -0.30$, $Z = -3.45$, $p < .001$), whereas in early retention, groups with reduced frequency outperformed the 100% frequency groups ($\Delta = 0.19$, $Z = 1.69$, $p < .05$). In late retention, significant population *ES* exist for groups performing a combination of program and parameter learning (performance: $\Delta = 0.52$, $Z = 2.90$, $p < .01$; consistency: $\Delta = 0.63$, $Z = 2.30$, $p < .01$). For isolated parameter learning we found no significant effects. Fail Safe N values range from 0 to 11 studies for low effect sizes (0.2). Sample error variance ranges from 27% to 100%.

Task Context. With respect to the context of the task we distinguish between simple laboratory tasks and gross, complex sport movements. We find significant effects exclusively for the laboratory context (performance measures). Acquisition performance of the reduced feedback groups is degraded (Δ = -0.25, Z = -1.75, p < .05) and early and late retention performance is enhanced (Δ = 0.27, Z = 2.15, p < .05 and Δ = 0.38, Z = 3.11, p < .001, respectively). For sport context and consistency measures we find no systematic pattern of results. Fail Safe N values ranging from 0 to 6 for a low effect are not satisfactory and sample error variance ranges from 48% to 100%.

Discussion

At first sight, the results of the global analysis (see Table 3) seem to confirm a reversal effect. That is, experimental groups which practice at a reduced frequency of augmented feedback show a lower performance level at the end of acquisition than experimental groups which apply 100% augmented feedback. In early as well as in late retention, this trend reverses, revealing experimental groups practicing with reduced augmented feedback to be the superior learners. This effect turns out to be even more obvious during late retention than during early retention. However, when scrutinizing the results, it becomes clear that – despite significant effect sizes for acquisition and late retention – the empirical evidence supporting the reversal effect is very limited. With a sum of 0.21 and 0.22 respectively, the effect sizes are reported as low, a situation which is further explained by the fact that the Fail Safe N values are also quite low: 2 and 3 respectively. This indicates that the inclusion of only two or three zero-effect studies which do not confirm the reversal effect would be enough to lower the effect size beneath the critical limit of 0.2. When considering the vote counting results (see Table 2), it is quite possible that such studies do indeed exist.

Because of the limited support for a reversal effect in the global analysis, it is important to evaluate the impact of moderator analysis, which might change the current interpretation of the reversal effect. One must be aware that the sample sizes (k) within some cases of moderator variables are very small. Therefore, the implications should be interpreted very cautiously.

The moderator analysis examined the differential effect of six moderators and examined a total of 20 factor levels. With only seven factor levels, the direction and intensity of the effects confirm the reversal effect (see Table 4). First, this is strong evidence for the reversal effect being more robust than suggested by the global analysis, but to a large extent also inconsistent. Additionally, one notices that in most cases the effect sizes decrease considerably from early to late retention. This can be taken as an indicator that differing performances in experimental groups with 100% augmented feedback compared to those with reduced augmented feedback might be more appropriately interpreted within the context of the specificity hypothesis than that of the guidance hypothesis. The specificity hypothesis (Schmidt, 1991) clearly targets the transfer effect. Differences in performance are explained by differing execution conditions in retention compared to acquisition. While subjects who have practiced with reduced augmented feedback are accustomed to augmented feedback withdrawal, subjects who have practiced with 100% augmented feedback are not. Augmented feedback becomes part of the task, in which its withdrawal can lead to an (at least short-term) impairment of performance. During late retention, existing differences in performance are then evened out. Therefore changes in effect sizes from acquisition to early retention are better attributed to a transfer effect than to a reversal effect.

The weakness and inconsistency of the reversal effect is further reflected by the finding that only when acquisition trials exceeding 90 is a mean effect size associated with an acceptable Fail Safe N. That is to say that if in acquisition more than 90 trials are done with reduced augmented feedback, better performance is shown in early and late retention compared to those realised with 100% augmented feedback. The guidance hypothesis points to the dependence on 100% augmented feedback and the maladaptative short-term correction is among other factors responsible for the reversal effect. These phenomena seem to be perceptible, or at least seem to occur only if acquisition is supported by a large number of trials. Even if the presented meta-analysis has not considered interactions, reference can be made to the results of the moderator variable "context of task". It seems perfectly plausible to say that disturbing effects, when learning with 100% augmented feedback, mainly occur when augmented feedback is offered frequently in simple artificial laboratory tasks. A distinct ceiling effect might possibly be assumed. There is every reason to believe that if one examines the changes in performance during acquistion in most of the studies in question (e.g., Goodwin & Meeuwsen, 1995), in all cases the final performance level of acquisition had already been reached after the first or second trial block in acquisition. Despite slight variations in the task, the subjects are offered 100% augmented feedback and it does not seem to be surprising that this promotes dependency and "over-reaction". The performance-degrading effect under reduced augmented feedback conditions in acquisition is low (Δ = 0.26) and – with a Fail Safe N of 2 – is almost negligible.

Overall, the moderator analysis makes clear that the reversal effect only occurs under very specific conditions. Therefore, it does not seem appropriate to generally speak of reduced augmented feedback being performance-degrading in acquisition and at the same time of its tendency to improve performance in retention. The findings from this meta-analysis also have implications for the practical application of reduced frequency effects, as well as for the contemporary account of reduced augmented feedback effects-evidence hypothesis. Considering the results concerning the task context (non-sport related vs. sport related tasks), the frequency of augmented feedback has no influence, especially on the learning of sport tasks. It is possible that the result-induced feedback in sport tasks is much more present and can be used in order to correct the movement without augmented feedback (Russell & Newell, 2007). This missing negative effect of reduced augmented feedback should be interpreted as "exoneration": It is not necessary to give feedback after every practice trial in acquisition. For most learning procedures, bandwidth augmented feedback may be of particular advantage in integrating positive and preventing negative guidance effects. According to the width of the band, missing augmented feedback acts as a reinforcement, and given augmented feedback leads to the desired movement.

The second issue concerns the current theoretical accounts for the assumed reversal effect. The phenomenon of the reversal effect is explained by way of ex post facto analysis, and the guidance hypothesis is the most widely used explanation. This hypothesis implies particular assumptions about information processing strategies (guidance, maladaptive short-term corrections; see Schmidt, 1991). Even though a current study of Anderson, Sekiya, Magill, and Ryan (2005) seems to provide evidence for one of the above mentioned assumptions of the reversal effect, the ambiguous impact of the guidance hypothesis is frequently revealed particularly when other than impoverished learning environments were used (Sherwood & Lee, 2003).

The former experimental designs include blocks of trials and measures of error based on means which are not appropriate for testing the assumptions of the guidance hypothesis. On the contrary, a trial-to-trial research method should be applied (Magill, 2001). Blackwell, Simmons, and Spray (1991) used this method to investigate the role of KR associated with the contextual interference effect, whilst Blackwell and Newell (1996) did the same to verify the persisting calibrating and temporary modulating effects of KR resp. No-KR. This strategy makes more sense than reproducing effects by varying the number of trials that should be in summary augmented feedback or the percentage of augmented feedback, and then interpreting these results ex post facto. A trial-to-trial analysis may be the best way to reveal immediate and delayed effects of augmented information, and to produce recommendations for practice that are based both on theory and solid evidence. In a first attempt to apply this method, Marschall, Müller, and Blischke (1997) could not confirm the assumptions of the guidance hypothesis concerning information processing strategies. More recently, Müller, and Blischke (2000) developed a simple model for using trial-to-trial analysis to quantify the extent to which different correction strategies result from different feedback conditions. These data indicated the first empirical studies (Brückner, Müller, Blischke, Shea, & Wright, 2001; Müller, Brückner, Panzer, & Blischke, 2001). This reveals that modifications and further research is needed to develop an appropriate method for investigating information processing strategies dependent on particular augmented feedback conditions.

References

(References marked by * were included in the meta-analysis)

Anderson, D. I., Sekiya, H., Magill, R. A., & Ryan, G. (2005). Support for an explanation of the guidance effect in motor skill learning. *Journal of Motor Behavior, 37*, 231-238.

Ashworth, S. D., Osburn, H. G., Callender, J. C., & Boyle, K. A. (1992). The effects of unrepresented studies on the robustness of validity generalization results. *Personal Psychology, 45*, 341-361.

*Behrman, A. L., Vander Linden, D. W., & Cauraugh, J. H. (1992). Relative frequency knowledge of results: Older adults learning a force-time modulation task. *Journal of Human Movement Studies, 23*, 233-250.

Blackwell, J. R., & Newell, K. M. (1996). The informational role of knowledge of results in motor learning. *Acta Psychologica 92*, 119-129.

Blackwell, J. R., Simmons, R. W., & Spray, J. A. (1991). Time series analysis of knowledge of results effects during motor skill acquisition. *Research Quarterly for Exercise and Sport, 62*, 10-17.

Bliesener, T. (1999). Methoden der systematischen Befundintegration – Eine Übersicht über Möglichkeiten und Grenzen der Meta-Analyse. [Methods of systematical research synthesis – review on chances and limitations of meta analysis] In J. Wiemeyer (Hrsg.), *Forschungsmethodologische Aspekte von Bewegung, Motorik und Training im Sport* (pp. 55-72). Hamburg, Germany: Czwalina.

*Blischke, K., Daugs, R., Igel, C., Panzer, S., & Müller, H. (1993). *Zum Einfluß von „Video-Fading"-Prozeduren auf das sportmotorische Lernen und Techniktraining* [Effects of „video-fading"-schedules on motor learning and technique training]. Unpublished manuscript, Universität des Saarlandes, Saarbrücken, Germany.

Blischke, K., Marschall, F., Müller, H., & Daugs, R. (1999). Augmented information in motor skill acquisition. In Y. Vanden Auweele, F. Bakker, S. Biddle, M. Durand, & R. Seiler (Eds.), *Psychology for physical educators* (pp. 257-287). Champaign/ IL.: Human Kinetics.

Bortz, J., Lienert, G. A., & Boehnke, K. (1990). *Verteilungsfreie Methoden der Biostatistik* [Non-parametric methods of biostatistics]. Berlin, Germany: Springer.

*Broker, J. P., Gregor, R. J., & Schmidt, R. A. (1993). Extrinsic feedback and the learning of kinetic patterns in cycling. *Journal of Applied Biomechanics, 9*, 111-123.

Brückner, S., Müller, H., Blischke, K., Shea, C., & Wright, D. (2001). A test of a simple model: Do different KR-schedules invoke different correction strategies? In J. Mester, J. King, H. Strüder, E. Tsolakis & A. Osterburg (Eds.), *European College of Sport Science: Perspectives and profiles: Book of abstracts of the 6th annual congress of the European College of Sport Science, 15th congress of the German Society of Sport Science* (p. 317). Köln: Sport und Buch Strauss.

*Butler, M. S., & Fischman, M. G. (1996). Effects of bandwidth feedback on delayed retention of a movement timing task. *Perceptual and Motor Skills, 82*, 527-530.

*Butler, M. S., & Fischman, M. G. (1999). Effects of performance-based schedules of feedback on the learning of a two-dimensional arm-positioning task. *Journal of Human Movement Studies, 36*, 183-201.

*Butler, M. S., Reeve, G. T., & Fischman, M. G. (1996). Effects of the instructional set in the bandwidth feedback paradigm on motor skill acquisition. *Research Quarterly for Exercise and Sport, 67*, 355-359.

*Carnahan, H., Vandervoort, A. A., & Swanson, L. R. (1996). The influence of summary knowledge of results and aging on motor learning. *Research Quarterly for Exercise and Sport, 67*, 280-287.

*Cauraugh, J. H., Chen, D., & Radlo, S. J. (1993). Effects of traditional and reversed bandwidth knowledge of results on

motor learning. *Research Quarterly for Exercise and Sport, 64*, 413-417.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
Drinkmann, A. (1990). *Methodenkritische Untersuchungen zur Metaanalyse* [Critical investigations of meta-analysis]. Weinheim, Germany: Deutscher Studienverlag.
Dunham, P. J. R., & Mueller, R. (1993). Effect of fading knowledge of results on acquisition, retention and transfer of a simple motor task. *Perceptual and Motor Skills, 77*, 1187 - 1192.
Field, A. P. (2003). The problems in using fixed-effects models in meta-analysis of real-world data. *Understanding Statistics, 2*, 105-124.
*Gable, C. D., Shea, C. H., & Wright, D. L. (1991). Summary knowledge of results. *Research Quarterly for Exercise and Sport, 62*, 285-292.
*Goodwin, J. E., & Meeuwsen, H. J. (1995). Using bandwidth knowledge of results to alter relative frequencies during motor skill acquisition. *Research Quarterly for Exercise and Sport, 66*, 99-104.
*Greuter, N. (1996). *Zum Einfluß der Frequenzverteilung biomechanischer Ergänzungsinformation auf die Aneignungs- und Behaltensleistungen einer sportmotorischen Bewegungsaufgabe* [Effects of biomechanical feedback frequency on acquisition and retention of a motor skill]. Unpublished diploma thesis. Universität des Saarlandes, Saarbrücken, Germany.
Guadagnoli, M. A., Dornier, L. A., & Tandy, R. D. (1996). Optimal length for summary knowledge of results: The influence of task-related experience and complexity. *Research Quarterly for Exercise and Sport, 67*, 239-248.
*Guay, M., Salmoni, A., & McIlwain, J. (1992). Summary knowledge of results for skill acquisition: Beyond Lavery and Schmidt. *Human Movement Science, 11*, 653-673.
Hagger, M. S. (2006). Meta-analysis in sport and exercise research: Review, recent developments, and recommendations. *European Journal of Sport Science, 6*, 103-115.
*Hebert, E. P., & Landin, D. (1994). Effects of a learning model and augmented feedback on tennis skill acquisition. *Research Quarterly for Exercise and Sport, 65*, 250-257.
Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), The handbook of research synthesis (pp. 285-300). New York: Sage.
*Hillebrecht, M. (1994). *Zum Einfluß der Informationsfrequenz auf das Erlernen von sportlichen Bewegungen* [Effects of information frequency on motor skill learning]. Frankfurt, Germany: Peter Lang.
*Hillebrecht, M., & Schuster, T. (1994). Individualisierung von Feedback bei motorischen Aufgaben [Individualization of feedback on motor tasks]. *Leipziger Sportwissenschaftliche Beiträge, 35*, 82-95.
*Jarus, T. (1995). Is more always better? Optimal amounts of feedback in learning to calibrate sensory awareness. *Occupational Therapy Journal of Research, 15*, 181-197.
Kernodle, M. W., & Carlton, L. G. (1992). Information feedback and the learning of multiple-degree-of-freedom activities. *Journal of Motor Behavior, 24*, 187-196.
Lai, Q., & Shea, C. H. (1999). The role of reduced frequency of knowledge of results during constant practice. *Research Quarterly for Exercise and Sports, 70*, 33-40.
*Lai, Q., Shea, C. H., Wulf, G., & Wright, D. L. (2000). Optimizing generalized motor program and parameter learning. *Research Quarterly for Exercise and Sport, 71*, 10-24.
*Lee, T. D., & Maray, B. K. V. (1994). Effects of bandwidth goals and bandwidth knowledge of results on motor learning. *Research Quarterly for Exercise and Sport, 65*, 244-249.
Magill, R. A. (2001). Augmented feedback in skill acquisition. In R. N. Singer, H. A. Hausenblas, & C. M. Janelle (Eds.), *Handbook of sport psychology* (2nd ed., pp. 86-114). New York: Wiley & Sons.
*Marschall, F. (1992). *Informationsfrequenz und motorisches Lernen* [Information frequency and motor learning]. Frankfurt, Germany: Peter Lang.
Marschall, F., Blischke, K., & Müller, H. (1997). Mit weniger externer Information besser lernen? Zum Problem der Beschreibung und Erklärung eines Phänomens [Better learning with less information? On the problem of the description and explanation of a phenomenon]. In P. Hirtz & F. Nüske (Eds.), *Bewegungskoordination und sportliche Leistung integrativ betrachtet* (pp. 241-245). Hamburg, Germany: Czwalina.
Marschall, F., Müller, H., & Blischke, K. (1997). Der "guidance"-Effekt beim motorischen Lernen: Zur empirischen Prüfung von Teilannahmen der "guidance"-Hypothese [The "guidance"-effect and motor learning: Examining the assumptions of the "guidance"-hypothesis]. In E. Christmann, J. Maxeiner & D. Peper (Eds.), *Psychologische Aspekte beim Lernen, Trainieren und Realisieren sportlicher Bewegungshandlungen* (pp. 88-94). Köln, Germany: bps.
Müller, H., & Blischke, K. (2000). Correcting errors in the presence of noise in the motor system: A simple model. *Journal of Sport and Exercise Psychology, 22*, 81.
Müller, H., Brückner, S., Panzer, S., & Blischke, K. (2001). Correcting errors in the presence of noise in the motor system: A formal model and first data. In J. Mester, J. King, H. Strüder, E. Tsolakis, & A. Osterburg (Eds.), *European College of Sport Science: Perspectives and profiles: Book of abstracts of the 6th annual congress of the European College of Sport Science, 15th congress of the German Society of Sport Science* (p. 129). Köln: Sport und Buch Strauss.
*McCullagh, P., & Little, W. S. (1990). Demonstrations and knowledge of results in motor skill acquisition. *Perceptual and Motor Skills, 71*, 735-742.
*Nicholson, D. E., & Schmidt, R. A. (1991). Scheduling information feedback to enhance training effectiveness. Proceedings of the Human Factors and Ergonomics Society, USA, 35, 1400-1402.
Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-322). New York: Sage.
Rosenthal, M. C. (1994). The fugitive literature. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 85- 96). New York: Sage.
Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York: Sage.
Rosenthal, R. & DiMatteo, M. R. (2001). Meta Analysis: Recent Developments in Quantitative Methods for Literature Reviews. *Annual Review of Psychology, 52*, 59-82.
Russel, D. M., & Newell, K. M. (2007). On No-KR tests in motor learning, retention and transfer. *Human Movement Science,26*, 155-173.
Rustenbach, J. S. (2003). *Metaanalyse. Eine anwendungsorientierte Einführung.* Bern, Switzerland: Hans Huber.
Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and perceptual motor learning. *Psychological Bulletin, 95*, 355-386.
*Schlicher, R. (1996). *Experimentelle Untersuchung zum Einfluß reduzierter und maximaler Sollwert-Präsentationsfrequenzen bei videogestützten Techniktrainings-Prozeduren im Sport* [On the effect of reduced and maximum presentation of goal information in video-assisted skill-training in sport]. Unpublished diploma thesis. Universität des Saarlandes, Saarbrücken, Germany.
Schlicht, W. (1994). *Sport und Primärprävention* [Sport and primary prevention]. Göttingen, Germany: Hogrefe.
Schmidt, R. A. (1982). *Motor control and learning. A behavioral emphasis.* Champaign/IL: Human Kinetics.
Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor neuroscience* (pp. 59-75). Dordrecht: Kluwer.
*Schmidt, R. A., Lange, C., & Young, D. E. (1990). Optimizing summary knowledge of results for skill learning. *Human Movement Science, 9* , 325-348.
*Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition. Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 352-359.
Schwarzer, R. (1997). *Meta-Analysis 5.3. The manual.* Retrieved September 28, 2005, from http://www.fu-berlin.de/gesund/gesu_engl/meta_e.htm.
Sherwood, D. E., & Lee, T. D. (2003). Schema Theory: Critical review and implications for the role of cognition in a new

theory of motor learning. *Research Quarterly for Exercise and Sport, 74*, 376-382.

*Sidaway, B., Moore, B., & Schoenfelder-Zohdi, B. (1991). Summary and frequency of KR presentation effects on retention of a motor skill. *Research Quarterly for Exercise and Sport, 62*, 27-32.

*Smith, P. K., Taylor, S. J., & Withers, K. (1997). Applying bandwidth feedback scheduling to a golf shot. *Research Quarterly for Exercise and Sport, 68*, 215-221.

*Sparrow, W. A., & Summers, J. J. (1992). Performance on trials without knowledge of results (KR) in reduced relative frequency presentations of KR. *Journal of Motor Behavior, 24*, 197-209.

Swinnen, S. (1996). Information feedback for motor skill learning: A review. In H. Zelaznik (Ed.), *Advances in motor learning and control* (pp. 37-66). Champaign/IL.: Human Kinetics.

Thomas, J. R., Lochbaum, M. R., Landers, D. M., & He, C. (1997). Planning significant and meaningful research in exercise science: Estimating sample size. *Research Quarterly for Exercise and Sport, 68*, 33-43.

Vickers, J. N. (1994). Psychological research in sport pedagogy. *Sport Science Review, 3*, 28-40.

*Weeks, D. L., & Sherwood, D. E. (1994). A comparison of knowledge of results scheduling methods for promoting motor skill acquisition and retention. *Research Quarterly for Exercise and Sport, 65*, 136-142.

*Wiemeyer, J. (1995). *Bewegungslernen im Sport. Motorische, kognitive und emotionale Aspekte* [Motor learning in sport: Motor, cognitive, and emotional aspects]. Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.

Wiemeyer, J. (1998). Schlecht üben, um gut zu lernen? Narrativer und meta-analytischer Überblick zum Kontext-Interferenz-Effekt [Bad practice for good learning? A narrative and meta-analytic overview of the contextual interference effect]. *Psychologie und Sport, 5*, 82-105.

*Winstein, C. J., Pohl, P. S., & Lewthwaite, R. (1994). Effects of physical guidance and knowledge of results on motor learning: Support for the guidance hypothesis. *Research Quarterly for Exercise and Sport, 65*, 316-323.

*Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 677-691.

*Wulf, G. (1992). Reducing knowledge of results can produce context effects in movements of the same class. *Journal of Human Movement Studies, 22*, 71-84.

*Wulf, G. (1994). *Zur Optimierung motorischer Lernprozesse* [On the optimisation of motor learning processes]. Schorndorf, Germany: Hofmann.

*Wulf, G., & Schmidt, R. A. (1989). The learning of generalized motor programs: Reducing the relative frequency of knowledge of results enhances memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 15*, 748-757.

*Wulf, G., Schmidt, R. A., & Deubel, H. (1993). Reduced feedback frequency enhances generalized motor program learning but not parameterization learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1-18.

We thank Robert N. Singer (University of Florida), David Wright (Texas A&M University) and Henry J. Rademacher (Universität des Saarlandes) for their critical notes and helpful grammatical and stylistic comments on former versions of this paper.

Original article received: 06.02.2007
Revised article received: 20.09.2007
Article accepted: 05.10.2007
Original article published: 05.11.2007