How can instructions and feedback with external focus be shaped to enhance motor learning in children? A systematic review

Ingrid P.A. van der Veer¹*, Evi Verbecque¹, Eugene A.A. Rameckers¹,²,³, Caroline H.G. Bastiaenen⁴¶, Katrijn Klingels⁴¶

¹ Rehabilitation Research Centre - REVAL, Faculty of Rehabilitation Sciences and Physiotherapy, Hasselt University, Hasselt, Belgium
² Department of Functioning and Rehabilitation, Research School CAPHRI, Maastricht University, Maastricht, the Netherlands
³ Centre of Expertise, Adelante Rehabilitation Centre, Valkenburg, the Netherlands
⁴ Department of Epidemiology, Functioning, Participation & Rehabilitation, Research School CAPHRI, Maastricht University, Maastricht, the Netherlands

* Corresponding author: Ingrid.vanderveer@uhasselt.be (lvdV)
¶ Joint senior authors

Abstract

Aim: This systematic review investigates the effectiveness of instructions and feedback with external focus applied with reduced frequency, self-controlled timing and/or in visual or auditory form, on the
performance of functional gross motor tasks in children aged 2 to 18 with typical or atypical development.

Methods: Four databases (PubMed, Web of Science, Scopus, Embase) were systematically searched (last updated May 31st 2021). Inclusion criteria were: 1. children aged 2 to 18 years old; 2. Instructions/feedback with external focus applied with reduced frequency, self-controlled timing, and/or visual or auditory form as intervention, to learn functional gross motor tasks; 3. Instructions/feedback with external focus applied with continuous frequency, instructor-controlled timing, and/or verbal form as control; 4. performance measure as outcome; 5. (randomized) controlled studies. Article selection and risk of bias assessment (with the Cochrane risk of bias tools) was conducted by two reviewers independently. Due to heterogeneity in study characteristics and incompleteness of the reported data, a best-evidence synthesis was performed.

Results: Thirteen studies of low methodological quality were included, investigating effectiveness of reduced frequencies (n = 8), self-controlled timing (n = 5) and visual form (n = 1) on motor performance of inexperienced typically (n = 348) and atypically (n = 195) developing children, for acquisition, retention and/or transfer. For accuracy, conflicting or no evidence was found for most comparisons, at most time points. However, there was moderate evidence that self-controlled feedback was most effective for retention, and limited evidence that visual analogy was most effective for retention and transfer. To improve quality of movement, there was limited evidence that continuous frequency was most effective for retention and transfer.

Conclusion: More methodologically sound studies are needed to draw conclusions about the preferred frequency, timing or form. However, we cautiously advise considering self-controlled feedback, visual instructions, and continuous frequency.

Registration: Prospero CRD42021225723
Key words: motor learning, child, adolescent, instruction, feedback, external focus

Introduction

Children apply many different gross motor skills in a wide variety of contexts, such as physical education (PE) classes, sports and playtime (1). These so-called functional skills are defined as motor skills used in sports or other daily life activities that entail relatively complex movement organization (2). Most children learn these skills almost effortlessly. Their increasing gross motor competence results from the interaction between factors in child (e.g. age, executive functions, psychological characteristics, and motor skill level), task (e.g. rules of the game, type of task, and level of task complexity) and environment (e.g. opportunities for PE and sports) (1,3–5). However, motor skills learning can be challenging for some children, due to neurological conditions (6,7) or neurodevelopmental disorders (8–11). Motor learning can be defined as a set of processes associated with practice or experience leading to relatively permanent improvements in the capability for producing motor skills (12). Instructors, like PE teachers, trainers, coaches, and occupational and physical therapists, apply motor learning on a daily basis (13–16). They use various motor learning variables, such as instructions and feedback, which they adapt to the child and the task practised (15–19). Their instructions and feedback are shaped by parameters, such as content (e.g. a specific focus of attention), frequency, form (e.g. visual or verbal), and timing (self- or instructor-controlled) (18,20,21).

With implicit motor learning, a child learns without awareness and with no or minimal increase in verbal knowledge (22). It is suggested that children benefit from this type of learning, because there is minimal involvement of the working memory (2,23,24). Implicit motor learning can be shaped by using an external focus of attention (EF) (23). With an EF, the child’s attention is steered to the impact of the movement on the environment (23). A recent systematic review investigated effectiveness of implicit
learning strategies in functional motor skills learning in typically developing children (TDC) (23). They concluded that the use of an EF appeared to be as, or even more, effective than an internal focus of attention (IF), which directs the child's attention to its body movements (23). An EF was also more effective than an IF in motor learning for children with Mild Intellectual Disabilities (MID) (25) and Attention Deficit and Hyperactivity Disorder (ADHD) (26). However, an IF appeared more effective in children with Autism Spectrum Disorder (ASD) (27). In children with Developmental Coordination Disorder (DCD), no differences were found for retention and transfer between groups using an EF or an IF (28,29).

When using an EF in practical settings, instructors have to decide how often (frequency), when (timing) and in what form to provide their instructions and feedback (20). Feedback can be provided after each trial (continuous frequency) or after a number of trials (reduced frequency) (30,31). It is indicated that reduced frequency is preferred in stroke patients (31). However, in (a)typically developing children, this remains unclear (30). The timing of instructions and feedback can be determined by the instructor (instructor-controlled) or the child (self-controlled) (32). Self-controlled feedback may enhance children’s motor learning more than instructor-controlled feedback (32). Most instructions and feedback are provided verbally (23,30,32) but instructors also use visual, tactile, and auditory (e.g. sound beeps) forms (14,17,19,20). Currently, it remains unclear what frequency, form and timing are to be preferred when using instructions and feedback with EF (14,30,32).

While previous reviews suggest that the effectiveness of EF may be moderated by child and task characteristics, like working memory capacity, motor skill level and type of task (23,32), we hypothesize that the effectiveness of EF may also be moderated by the instructors’ chosen frequency, timing, and form. Therefore, this systematic review investigates the effectiveness of instructions and feedback with EF applied with reduced frequency, in visual or auditory forms, and/or on request of the child (I), compared to instructions and feedback with EF applied with continuous frequency, in verbal form,
and/or initiated by the instructor (C), on the performance of functional gross motor tasks (O) in children aged 2 to 18 with typical and atypical development (P).

Methods

A systematic review of randomized controlled trials (RCTs) and non-randomized controlled clinical trials (CCTs) was performed. The hypotheses were: 1. instructions and feedback with EF applied with reduced frequency will be more effective than those applied with continuous frequency; 2. self-controlled instructions and feedback with EF will be more effective than instructor-controlled instructions and feedback; and 3. visual or auditory instructions and feedback with EF will be more effective than verbal instructions and feedback. This systematic review is written according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses 2020 (PRISMA 2020) (33,34) and registered in the international prospective register of systematic reviews (PROSPERO) under registration number: CRD42021225723.

Inclusion and exclusion criteria

Inclusion and exclusion criteria were defined in line with the PICOT structure (Population, Intervention, Control, Outcome, Type of study).

Inclusion criteria were:

1. Population: Children with (a)typical development aged 2-18 years. Studies which included a combined population of adolescents and adults were included if there were sub-analyses with adolescents.
2. Intervention: Instructions or feedback with EF applied with reduced frequency, in visual or auditory form and/or with self-controlled timing, used to learn functional gross motor tasks. An analogy, a metaphor that integrates the complex structure of the to-be-learned task (35), is considered an EF because a child aims to reproduce the metaphor (36). Reduced frequencies can be applied in fixed frequency (feedback after a fixed number of trials) or faded frequency (reducing the frequency over time) (30,31).

3. Control: Instructions and feedback with EF applied with continuous frequency, in verbal form and/or with instructor-controlled timing.

4. Outcome: A performance measure (e.g. accuracy or quality of movement) as primary outcome, used to assess acquisition and/or learning of functional gross motor tasks. Acquisition is measured during practice blocks or with a post-intervention test (“post-test”), and learning is measured with retention and/or transfer tests (37).

5. Type of study: Studies using a RCT or CCT without randomization design.

6. Publication type: Publications of original RCTs and CCTs.

7. Language: Studies written in English or Dutch.

Exclusion criteria were:

1. Population: Children with (a)typical development under the age of 2 years or adults.

2. Intervention: Intervention methods like Neuromotor Task Training, because they provide no insight into effectiveness of separate instructions or feedback; instructions and feedback used to learn laboratory, fine motor and static balance tasks, because they did not meet the definition of functional gross motor task (2).

3. Control: A tactile form of instructions and feedback, because of its IF.
4. Outcome: Outcome measures that assessed brain anatomy and functions as primary outcomes.

5. Type of study: Studies performed with designs other than RCT and non-randomized CCT.

6. Publication type: Conference proceedings/reports and books.

7. Language: Studies not written in English or Dutch.

**Literature search**

A systematic search was conducted in PubMed, Web of Science, Scopus and Embase. The search was last updated on the 31st of May 2021. Because instructions and feedback are also used when applying practice conditions, a broad search query was used to ensure that no relevant studies were missed. The search terms concerned four key topics: motor learning, instruction, feedback, and practice conditions. These topics were combined as motor learning AND (instruction OR feedback OR practice conditions). An explorative search to inventory relevant search terms showed that, in title and abstract, participants were often described in general (e.g. subjects). It also showed that various outcome measures were used to assess motor task performance (e.g. accuracy, speed, count, distance). To prevent studies being missed, search terms did not incorporate terms related to population or outcome. No date restrictions or filters were applied. See S1 file for the detailed search queries.

**Study selection**

The eligibility of the studies was assessed in two phases: on title and abstract (phase 1); on full text (phase 2). The selection criteria were applied in a fixed sequence (population, intervention, control, outcome, type of study, publication type and language) by two reviewers independently (IvdV and EV). If necessary, authors were contacted for full texts. After each phase, a consensus meeting discussed the results of the article selection. Full text versions were read in case of disagreement after phase 1 and an
independent reviewer (ER) was consulted in case of disagreement after phase 2. References of the included studies and of the three systematic reviews concerning children’s motor learning \((23,30,32)\) were checked by one reviewer (IvdV) to ensure that all relevant studies had been included.

### Data extraction

Data were extracted using a standardized sheet by one reviewer (IvdV or EV) and checked and complemented by the other. Corrections and additions were discussed between both reviewers; in the case of disagreement, an independent reviewer (ER) was consulted. Authors were not contacted for further details about studies.

For each study, the following data were extracted: 1. Characteristics of the study design: information regarding the group allocation of the participants (e.g. randomization procedure), blinding of participants, assessors, outcome measures and all relevant data for analyses; 2. Population characteristics: number of participants in total and per group, age range, mean age and standard deviations (SD), skill level (inexperienced or trained), and diagnosis, if given; 3. Intervention characteristics: details about instructions or feedback to the experimental and control group(s), the task, and the practice sessions (e.g. frequency, volume and duration); 4. Outcome and assessment time points: the primary and secondary outcome(s) to measure motor performance and type and timing of measurements in acquisition and test phase (pre-, post-, retention and/or transfer tests); 5. Results: summary statistics with measures of precision for each group, the data for differences between groups, and thresholds of minimal clinically important differences.
Methodological quality assessment

The revised Risk of Bias tool (RoB2), for randomized trials (38), and the Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) (39), were used to assess methodological quality.

The RoB2 evaluates five major domains of biases: selection, performance, detection, attrition, and reporting biases. Signalling questions were answered to reach a domain-specific RoB judgement of ‘low’, ‘some concerns’ or ‘high’. If not referred to a registered trial protocol, Questions 5.2 and 5.3 were answered based on the data-analysis section. Using the judgements of the five domains, an overall RoB judgement was made. If at least four domains were of some concern, the overall RoB was considered high.

The ROBINS-I evaluates seven major domains of biases: confounding, selection, classification, performance, detection, attrition, and reporting biases. As for the RoB2, signalling questions were used to reach a domain-specific RoB judgement of ‘low’, ‘moderate’, ‘serious’, ‘critical’ or ‘no information’. If not referred to a registered trial protocol, Questions 7.1, 7.2 and 7.3 were answered based on the data-analysis section. Based on the domain-specific judgements, an overall RoB judgement was made.

Four reviewers (IvdV, EV, ER and KK) investigated RoB. Each study was assessed by two reviewers independently. Consensus was reached in a meeting with all reviewers.

Analyses

Results were described for study selection, study characteristics and methodological quality. To answer the hypotheses, as a first step a meta-analysis was planned with studies comparable for study design, instructions and feedback, and task. Therefore, the instructions and feedback were coded according to each parameter (frequency, timing and form). For frequency, the intervention was coded as reduced
fixed or reduced faded frequency and the control as continuous frequency (hypothesis 1). For timing, the intervention was coded as self-controlled and the control as instructor-controlled (hypothesis 2). For form, the intervention was coded as visual or auditory and the control as verbal (hypothesis 3). Studies were grouped according to the type of comparison between coded intervention and control. Each task is defined by its own constraints, which are related to the context in which the task is performed (40). Only studies with similar tasks could be combined in a meta-analysis. After subgrouping in subsequent steps according to (firstly) task and (secondly) population (TDC and per diagnosis), it was still not possible to pool data due to heterogeneity and to the incompleteness of the reported data. Therefore, a best-evidence synthesis was performed. The best-evidence synthesis table was structured according to the parameter of interest (frequency, timing, or form) and subdivided into comparisons of coded interventions and controls, as described above. If studies included more than one group with reduced frequency, the frequency that was most comparable with other studies was used for analysis. Within comparisons, studies were ordered according to comparable tasks, mentioning studies of good methodological quality first to increase the prominence of the most trustworthy evidence. Results were described per outcome measure. The results of each study were rated as significant (favouring a specific frequency, timing or form), inconsistent or not significant (41). Then, the evidence for each comparison was rated according to the guidelines of van Tulder et al. (41): strong (consistent findings among multiple high quality RCTs), moderate (consistent findings among multiple low quality RCTs and/or CCTs and/or one high quality RCT), limited (one low quality RCT and/or CCT), conflicting (inconsistent findings among multiple RCTs and/or CCTs), or no evidence from trials (no RCTs or CCTs). Consistency was defined as 75% of the studies assessing the same comparison showing results in the same direction.
Results

Study selection

The search resulted in 3813 unique hits. After screening title and abstract, 3521 hits were excluded. The remaining 292 hits were screened on full text, eight of which met the inclusion criteria. Reasons for exclusion were not meeting the criteria for: population (n = 150), intervention (n = 84), control (n = 1), type of study (n = 41), publication type (n = 7) or language (n = 1). Of the excluded studies, 24 investigated effectiveness of instructions and feedback with EF in children’s functional gross motor learning in comparison with an IF and/or no instructions or feedback, without distinction in frequency, timing or form between groups (26–28, 42–62). Of the studies that distinguished in frequency, timing or form between groups, eight used an IF (63–70). One study was excluded because its control group also used reduced instead of continuous frequency (71) (S2 file: overview of the excluded studies that nearly met inclusion criteria). Additionally, five studies were found through the references check, resulting in a total of 13 included studies (Fig 1).

Study characteristics

Seven out of 13 studies included 348 inexperienced TDC (72–78), ages ranging from 6 (76,77) to 13 years (78). Seven studies included 195 inexperienced children with motor disabilities (77,79–84), ages ranging from 6 (77,80,83) to 18 years (80). Mean ages and SDs were not reported in five studies (75–77,79,84). The children with motor disabilities comprised children with MID (82), DCD (84), ASD (77,83) or CP (79–81). Overall, the studies involved small sample sizes, the number of participants per group ranging from 6 (84) to 16 (82), with six studies having samples of 10 or less (73,77,79–81,84). All studies used object
control tasks (72–84), 12 throwing (72–81,83,84) and one golf-putting (82). In 10 studies, participants practised only once (72–78,81–83), the number of trials ranging from 30 (75) to 90 (78,83). Participants in the remaining studies practised five times with a total of 100 trials (80), or eight times with a total of 240 trials (79,84) (Table 1).

The effectiveness of feedback with EF applied in reduced frequency compared to continuous frequency was investigated in eight studies (73–79,82), six of which included TDC (73–78). The remaining studies included children with ASD (77) or CP (79). The reduced frequency was applied in three fixed frequencies of 20% (82), 33% (74,75) and 50% (76–79), and one faded frequency decreasing from 100% to 0% with an average of 62% (73). All studies assessed accuracy (73–79,82), with two also measuring variability (74,78), and one quality of movement (75). Acquisition was assessed in all studies (73–79,82), while retention tests were used in seven (73–77,79,82), in which timing varied from 24 hours (73,75–77,82) to 1 week (74,82). Only three studies measured transfer (74,75,78), in which timing varied from immediately after practice (0 hours) (74,75,78) to 1 week (74) (Table 1).

Effectiveness of self-controlled feedback compared to instructor-controlled feedback to improve accuracy in object control tasks was investigated in five studies (72,73,80,81,84). TDC were included in two studies (72,73), while the others included children with DCD (84) or CP (80,81). In four studies, the frequency of the self- and instructor-controlled feedback was the same (72,80,81,84), while in one frequencies were different, 33% in the self-controlled group and 100% in the instructor-controlled group (73). All studies measured acquisition and retention (72,73,80,81,84). In most studies, retention was measured after 24 hours (72,73,80,81), though in one the timing was unclear (84). One-day transfer tests were used in two studies (Table 1).
One study with children with ASD and MID investigated the effectiveness of visual analogy compared to verbal analogy for improving accuracy in basketball shooting on acquisition, retention (24 hours), and transfer (0 and 24 hours) (83) (Table 1).

**Methodological quality**

Twelve RCTs were assessed with the RoB2, all of which having an overall RoB judgement of high (72–78,80–84) (Fig 2a). Although studies mentioned randomized groups, none described the generation method used and whether allocation was concealed (72–78,80–84). Only one study provided a demographic characteristics table (83). Most studies were at high risk for performance bias, none of the studies reported using intention-to-treat (ITT) analysis and how they handled missing data (72–78,80–84). Most studies were also at high risk for detection bias, only one study reported no missing data (84). In six studies, the F statistics showed that there were missing data, but information on the amount, at which time point and in which group was lacking (73–75,78,80,81). In most studies, outcome assessors were not blinded or it remained unclear whether they were blinded (72–82,84). None referred to a registered trial protocol, raising concerns about possible reporting bias (72–84). The study of Hemayattalab & Rostami (2010) (79) was the only non-randomized CCT included. It had an overall judgement of serious RoB due to a serious RoB in measurement of outcomes, while the remaining domains were at low RoB (79) (Fig 2b).

Best-evidence synthesis

Regarding frequency of feedback (hypothesis 1), three out of seven studies investigated the effectiveness of reduced fixed frequency in similar tasks (74,76,77). However, one reported no summary
statistics (74) and the other two had the same first author (76,77). The remaining studies used non-comparable tasks (75,78,79,82). Only one study examined the effectiveness of reduced faded frequency. As regards timing of feedback (hypothesis 2), four out of five studies included similar tasks (72,73,81,84), but summary statistics were lacking in two of these (72,73); the remainder included different populations (81,84), and only one investigated a visual form of instruction (hypothesis 3). Therefore, all studies were included in the best-evidence synthesis (72,73,82–84,74–81) (Table 2). Although each study described whether there were significant group differences, none mentioned thresholds for minimal clinically important differences (72,73,82–84,74–81).

The following paragraphs describe the results from the best-evidence synthesis for the parameters frequency, timing and form. For frequency, results were reported for the outcomes accuracy, variability and quality of movement. Studies of timing and form only assessed accuracy. For each parameter, results are ordered according to the following time points: 1. Acquisition measured during practice; 2. Acquisition measured with a post-test; 3. Retention; and 4. Transfer.

**Insert table 2 about here**

**Frequency**

The evidence whether reduced fixed frequency of feedback was more effective than continuous frequency (hypothesis 1) in improving accuracy of object control tasks on acquisition was conflicting (74–76,78,79,82). For acquisition measured during practice, continuous frequency appeared more effective in TDC and in children with ASD (77) or MID (82) (76,77); however, two other studies with TDC found no significant group differences (74,78). For acquisition measured with a post-test, the results of the studies varied with the population. No significant group differences were found in TDC (74,75), while continuous frequency appeared more effective in children with CP (79). For retention, conflicting evidence was also found (74–76,78,79,82): for TDC, two studies found no significant group differences (74,75), while two
other studies indicated that reduced frequency was more effective (76,77); for children with motor
disabilities, results showed that children with CP (79) and MID (82) performed best with reduced
frequency while children with ASD did best with continuous frequency (77). For transfer, no evidence
supported either frequency in TDC (74,75,78) (Table 2). Only one study compared reduced faded
frequency to continuous frequency to improve accuracy in beanbag throwing in TDC (73). For acquisition
measured during practice, they found no significant group differences (73). For retention, limited
evidence was found favouring continuous frequency (73) (Table 2).

There was no evidence that reduced fixed or continuous frequency was more effective in reducing
variability in throwing in TDC at all four time points (74,78). The study that assessed quality of
movement in soccer ball throwing in TDC found no evidence supporting either frequency for acquisition
measured with a post-test (75). However, for retention and transfer limited evidence was found
favouring continuous frequency (75). (Table 2).

Timing

For accuracy in object control tasks, conflicting evidence was found on effectiveness of self-controlled
versus instructor-controlled feedback (hypothesis 2) with equal frequency for acquisition measured
during practice (72,80,81). Of the studies including children with CP (80,81), one showed that self-
controlled timing was more effective (80), while another found no significant group differences (81); no
significant group differences were found in TDC (72). Also, no significant group differences were found in
children with DCD for acquisition measured with a post-test (84). For retention, the self-controlled group
performed best in three studies (72,81,84), including TDC (72), children with CP (81) and DCD (84). A
fourth study showed no significant group differences in children with CP (80), which resulted in only
moderate evidence favouring self-controlled timing (72,80,81,84). For transfer, the evidence was
conflicting in children with CP: while one study showed that self-controlled timing was more effective, another found no significant group differences (80,81) (Table 2).

One study used different frequencies in the self- and instructor-controlled groups to improve accuracy in beanbag throwing in TDC (73). For acquisition measured during practice, no evidence supported either timing. However, there was limited evidence that 100% instructor-controlled feedback was more effective than 33% self-controlled feedback for retention (73) (Table 2).

Form

One study investigated the effectiveness of visual analogy compared to verbal analogy (hypothesis 3) used to improve accuracy in basketball throwing in children with ASD and MID (83). For acquisition measured with a post-test, no evidence supported either form (83). However, for retention limited evidence was found favouring a visual form of instruction (83) (Table 2).

Discussion

The aim of this systematic review was to investigate the effectiveness of instructions and feedback with EF applied with reduced frequency, with self-controlled timing or in visual form in the learning by (a)typically developing children of functional gross motor tasks. Previous research investigating effectiveness of instructions or feedback with EF found conflicting results for children (23,32) and adults (37,85). It was hypothesized that the frequency, timing and/or form of instructions and feedback (20) influenced their effectiveness. The following paragraphs will discuss results by each hypothesis.

First, it was hypothesized that reduced frequency would be more effective than continuous frequency. However, the results of the best-evidence synthesis did not support this. On the contrary, limited evidence showed that, for retention, continuous frequency was more effective than reduced faded
frequency to improve accuracy in throwing beanbags in TDC (73). Also, limited evidence favoured continuous frequency to improve quality of movement in soccer ball throwing in TDC for retention and transfer (75) (Table 2). For acquisition, conflicting evidence was found, but studies found either no significant group differences (73–75,78) or significant differences favouring continuous frequency (76,77,79,82). A possible reason why continuous frequency appeared more effective could be the short practice duration, as most studies included only one practice session (73–78,82) (Table 1). At the beginning of the learning process, more information (e.g. by means of more instructions and feedback) is needed to acquire new skills (12,86,87). With inexperienced children, it is likely that some children remained in the early learning stage due to insufficient repetitions and, therefore, performed better with continuous frequency. In practical settings, children have longer training periods. Therefore, future studies adopting longer practice durations would be more of more practical interest. This limited or conflicting evidence is in line with previous research. A systematic review investigating effectiveness of frequency of feedback to improve upper limb motor skills in TDC and children with CP found limited evidence that, for TDC, continuous frequency was more effective for acquisition and reduced faded frequency for retention when learning laboratory tasks (30). For reduced fixed frequencies in TDC, no conclusions could be drawn due to inconsistency in results (30). Based on three pre-post design studies, the review concluded that both faded and continuous frequency improved upper limb motor learning (30). As such, it was not possible to draw conclusions about the preferred frequency.

Secondly, it was hypothesized that self-controlled timing would be more effective than instructor-controlled timing. The results of the best-evidence synthesis confirmed this, with moderate evidence for retention when frequency of feedback was similar in both groups (72,80,81,84) (Table 2). On the contrary, when frequencies were dissimilar, the instructor-controlled group appeared more effective for retention (73). This inconsistency may be due to the frequency of feedback, as the self-controlled group received less feedback than the instructor-controlled group (Table 1) (73). For all other time points,
either no or conflicting evidence was found. This might be due to the low methodological quality of the included studies, which will be elaborated later. A systematic review investigating the effectiveness of autonomy support in children's functional skill motor learning yielded similar results (32). It found that self-controlled feedback was more effective in several studies, but it was argued that child characteristics, like trait anxiety, cognitive skills and age, may have influenced effectiveness (32). The advantages of self-controlled feedback were also found in adults (36). However, both reviews included studies with EF and IF (32,36). Although more evidence is needed to draw conclusions for all time points, the results from the best-evidence synthesis, supported by previous research, suggests that instructors should consider using self-controlled timing in children's motor learning.

Finally, it was hypothesized that children learnt functional gross motor skills best with a visual form of instructions and feedback compared to a verbal form. However, only one study investigated this specific comparison (83). Post-hoc comparisons showed that children with ASD threw more accurately after a visual analogy (83). Similar results were found in studies with healthy young adults and young adults with Down syndrome, where skill performance improved more after video (88,89) or instructor demonstration (90) than with verbal instructions with EF. Although evidence is limited, instructors might consider using pictures, videos or real live demonstrations as instructions or feedback to teach children motor skills.

This was the first study to systematically investigate the modifying role of frequency, timing and form in instructions and feedback with EF on children's motor learning. A strength of this study was that it followed a registered protocol, comprising a selection process and RoB assessment performed by two reviewers independently, with a third to be consulted in cases of disagreement. Furthermore, RoB was assessed by means of reference standards (the Cochrane RoB tools) and findings were analysed according to a prespecified plan. There was no need to contact authors of included studies for further
details. There is a small possibility that we interpreted reported information slightly different than meant by the authors.

Providing recommendations for instructors about the frequency, timing and form of instructions and feedback with EF appeared challenging for two particular reasons. Firstly, drawing evidence-based conclusions was difficult because of the poor methodological quality of the studies (72,73,82–84,74–81) (Fig. 2). In particular, blinding of outcome assessors, analysing according to ITT, and handling missing data properly require attention in future studies (91,92). Furthermore, authors should report methods and results in more detail, essential for adequately determining the RoB (91,92). It is possible that methodological quality appeared lower due to insufficient reporting of details. Additionally, the generally small sample sizes and the lack of reported thresholds of clinically meaningful differences also hindered interpretation. Inadequate sample sizes increase the risk of finding non-significant results or contrary conclusions with similar studies (93,94). This might have influenced the number of non-significant results found in individual studies and, more specifically, the lack of evidence or the conflicting evidence in the best-evidence synthesis (Table 2) (93,95). In particular, the results of the post-hoc comparisons should be interpreted cautiously (93). Although some studies found significant differences, it remains unclear whether these differences are large enough to be relevant in practical settings (96,97). More methodologically sound studies based on proper sample size calculations are needed to draw conclusions regarding the preferred frequency, timing and form of instructions and feedback.

Secondly, generalizability of the results was hampered because all included studies used object control tasks with inexperienced children, and measured accuracy. This overrepresentation of tasks, skill level and outcome is in line with previous research (23,32). In therapy, PE classes and sports, children learn various tasks with different levels of complexity (98) and, depending the child’s needs, instructors teach new tasks to novice children or optimize existing skills in experienced or trained children (8,99,100). The challenge point framework conceptualizes the amount and specificity of information needed to
learn skills, based on the level of task complexity, the skill level of the individual, and the interaction of level of complexity with skill level (86). Thus, instructors should adapt frequency, timing and form of instructions and feedback to the individual and the task. Future research should attempt to include a wider variety of tasks and/or skill levels in studies, firstly, because this will improve generalizability, and secondly, to gain a better understanding of the influencing roles of task and skill level on the effectiveness of frequency, timing and form. In order to guarantee comparability of studies, a framework that classifies tasks based on their characteristics could be helpful. Future research should give attention to developing such a framework. Potentially relevant characteristics are the number of degrees of freedom, cognitive demands, sequence of movement structure, spatial and temporal demands, and the context of tasks (2,40,87). As for outcome, few studies assessed variability (74,78) or quality of movement (75), as well as accuracy. In practical settings, instructors often focus on improving functionality instead of normality (8,99,100). From that point of view, accuracy is a relevant outcome, because it focuses on the result of the performance instead of on the optimal movement pattern. However, instructors can target various improvements, depending on the child’s need. Therefore, more result-related outcomes (e.g. variability, number of successful attempts and distance) and movement pattern-related outcomes (e.g. quality of movement and kinematic variables) should be considered in future studies. Irrespective of the chosen outcome, researchers should use valid, reliable and responsive outcome measures.

**Conclusion**

Based on the results of this systematic review, instructors should consider using self-controlled feedback with EF to enhance children’s motor learning (moderate evidence). Regarding a specific frequency or form, no conclusions can be drawn yet. However, based on limited evidence, instructors could consider
using visual instructions and continuous frequency when aiming to improve quality of movement. Because specific child and task characteristics can also moderate the effectiveness of instructions and feedback (23,32,86), instructors should explore the optimal frequency, timing and form for each child until more research provides us with a better understanding of their moderating role. Future research should put effort into developing a framework that classifies tasks based on their characteristics. Furthermore, it should aim to advance insights into the modifying role of frequency, timing and form in instructions and feedback with EF with methodologically sound studies focusing on a variety of tasks, skill levels and outcome measures.

Acknowledgements

Not applicable.

References


16. Otte FW, Davids K, Millar SK, Klatt S. When and how to provide feedback and instructions to athletes?—How sport psychology and pedagogy insights can improve coaching interventions to enhance self-regulation in training. Front Psychol. 2020;11: 1444.


20. Levac D, Missiuna C, Wishart L, DeMatteo C, Wright V. Documenting the content of physical


expectedancies, positive affect, and motor learning. Psychol Sport Exerc. 2017;31: 28–34.


76. Zamani MH, Zarghami M. Effects of frequency of feedback on the learning of motor skill in...


Supporting information

S1 file. Search queries for the individual databases

S2 file. Excluded studies that nearly met inclusion criteria
Funding

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Data and materials availability statement

All relevant data are available in this manuscript or its Supporting Information files.

Authors’ contribution

IvdV, EV, ER, CB and KK designed the overall setup of the systematic review. IvdV and EV conducted the entire search, study selection and data extraction. Methodological quality was assessed by IvdV, EV, ER and KK. All authors drafted the manuscript and performed multiple revisions. All authors read and approved the final manuscript.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.
Fig 1. Prisma flow diagram of the study selection

Fig 1 Legend: n = number

Identification of studies via databases and registers

Records identified from:
PubMed (n = 1712)
Web of Science (n = 2631)
Scopus (n = 3079)
Embase (n = 2277)

Records removed before screening:
Duplicate records removed (n = 5885)

Records screened (n = 3813)

Records excluded (n = 3512)

Reports sought for retrieval (n = 292)

Reports not retrieved (n = 0)

Reports assessed for eligibility (n = 202)

Reports excluded:
Population (n = 150)
Intervention (n = 84)
Control (n = 1)
Type of study (n = 41)
Publication type (n = 7)
Language (n = 1)

Studies included in review (n = 13)
Reports of included studies (n = 13)

Reports identified through reference check (n = 6)
**Fig 2. Methodological quality of the included studies**

**Fig 2a. Methodological quality assessed with RoB2**

Fig 2a legend: D1 = selection bias; D2 = performance bias; D3 = detection bias; D4 = attrition bias; D5 = reporting bias; + = low risk; – = some concerns; × = high risk.

<table>
<thead>
<tr>
<th>Study</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiviacowsky et al. 2008</td>
<td>×</td>
<td>×</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>De Oliveira et al. 2009</td>
<td>–</td>
<td>×</td>
<td>×</td>
<td>–</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Gillespie 2003</td>
<td>–</td>
<td>×</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Hemayattalab 2014</td>
<td>–</td>
<td>×</td>
<td>×</td>
<td>+</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Hemayattalab et al. 2013</td>
<td>–</td>
<td>×</td>
<td>×</td>
<td>–</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Sabzi et al. 2012</td>
<td>–</td>
<td>×</td>
<td>×</td>
<td>+</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Sidaway et al. 2012</td>
<td>–</td>
<td>×</td>
<td>×</td>
<td>–</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Tse &amp; Masters 2019</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Wulf et al. 2010</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>+</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Zamani &amp; Zarghami 2015</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Zamani et al. 2015 (ASD/TDC)</td>
<td>–</td>
<td>×</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>×</td>
</tr>
<tr>
<td>Zamani et al. 2015 (DCD)</td>
<td>–</td>
<td>×</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>×</td>
</tr>
</tbody>
</table>

**Fig 2b. Methodological quality assessed with ROBINS-I**

Fig 2b legend: D1 = bias due to confounding; D2 = selection bias; D3 = classification bias, D4 = bias due to deviation from intended interventions; D5 = bias due to missing data; D6 = bias in measurement of outcomes; D7 = reporting bias; + = low risk; × = serious risk.
<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemayatlab &amp; Rostami 2010</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>+</td>
<td>X</td>
</tr>
<tr>
<td>Author(s), year</td>
<td>Population characteristics</td>
<td>Study design</td>
<td>Intervention characteristics</td>
<td>Study design</td>
<td>Intervention characteristics</td>
<td>Measurements</td>
<td>Outcome</td>
<td>Co-varyables</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>-----------------------------</td>
<td>--------------</td>
<td>-----------------------------</td>
<td>--------------</td>
<td>---------</td>
<td>--------------</td>
</tr>
<tr>
<td>Gillespie, 2003</td>
<td>Children with MID, IQ range 55-70 (n = 32)</td>
<td>RCT</td>
<td>• 20% frequency: KR after every fifth trial</td>
<td>• 100% frequency: KR after every trial</td>
<td>Golf-putting task</td>
<td>1 practice session of 10 blocks of 5 trials [total 50 trials]</td>
<td>Accuracy: score varied from 0-5 based on zones around the target (higher is better)</td>
<td>During acquisition:</td>
</tr>
<tr>
<td></td>
<td>Mean age in years (50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24h</td>
</tr>
<tr>
<td></td>
<td>Skill level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retention:</td>
</tr>
<tr>
<td>Hemayattalab &amp;</td>
<td>Children with CP, GMFCS 1 (n = 24)</td>
<td>Non-randomized CCT</td>
<td>• 50% frequency: KR on half of the trials (n = 8)</td>
<td>• 100% frequency: KR after every trial (n = 8)</td>
<td>Dart throwing</td>
<td>8 practice session of 6 blocks of 5 trials, period not reported [total 240 trials]</td>
<td>Accuracy: registered in points based on zones around the target (higher is better)</td>
<td>Pre-test</td>
</tr>
<tr>
<td>Postami, 2010</td>
<td>7-15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>During acquisition:</td>
</tr>
<tr>
<td></td>
<td>Inexperienced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retention:</td>
</tr>
</tbody>
</table>

*Significantly better than **no significant differences
De Oliveira and al., 2009

RCT

• Simple task + 25% frequency: KR after a quarter of the trials (n = 15)

• Simple task + 50% frequency: KR after half of the trials (n = 15)

• Simple task + 75% frequency: KR after three quarters of the trials (n = 15)

• Simple task + 100% frequency: KR after every trial (n = 15)

• Complex task + 25% frequency: KR after a quarter of the trials (n = 15)

• Complex task + 50% frequency: KR after half of the trials (n = 15)

• Complex task + 75% frequency: KR after three quarters of the trials (n = 15)

• Complex task + 100% frequency: KR after every trial (n = 15)

Bocha game throwing

• Simple task: throw with backward-forward pendulum movement of the arm

• Complex task: throw with same pendulum movement followed by an overhead circular movement of the arm

1 practice session of 9 blocks of 10 trials (total 90 trials)

Outcome

Accuracy: absolute error in cm (lower is better)

Variability: variable error in cm (lower is better)

During acquisition

A NOVA analysis with post-hoc Tukey HSD tests

Accuracy

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

Variability

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

Transfer:

High fidelity

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency

• Significant main effect for frequency
<table>
<thead>
<tr>
<th>Study</th>
<th>TIDC (n = 40)</th>
<th>RCT</th>
<th>Outcome</th>
<th>During acquisition</th>
<th>ANOVA analysis with post-hoc Tukey HSD tests</th>
<th>Variable (please specify)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salzi et al., 2012</td>
<td>10.4 (1.0)</td>
<td>Reduced frequency: KR with faded frequency from 100% to 0% (n = 10)</td>
<td>Accuracy: score varied from 0-100 based on zones around the target (higher is better)</td>
<td>Retention: 24h</td>
<td>Acquisition: No significant main effect for frequency</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100% frequency: KR after every trial (n = 10)</td>
<td></td>
<td></td>
<td>Retention: Significant main effect for groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self-controlled: KR on request in 3 out of every 10 trials (n = 10)</td>
<td></td>
<td></td>
<td>Post-hoc testing: 100% &gt; reduced feedback, SC and bandwidth</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bandwidth: feedback when score is less than 50 out of 100 points (n = 10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 practice session of 6 blocks of 10 trials (total 60 trials)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Throw with beanbag at target on the floor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIDWAY et al., 2012</td>
<td>10.7 (0.6)</td>
<td>Simple task + 33% frequency: KR after every third trial</td>
<td>Accuracy: absolute error based on the standard deviation about the mean score (lower is better)</td>
<td>Retention 0h and 1w</td>
<td>Acquisition: No significant main effect for frequency</td>
<td>Retention 0h and 1w: No significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simple task + 100% frequency: KR after every trial</td>
<td></td>
<td></td>
<td>Transfer 0h and 1w: No significant main effect for frequency</td>
<td>Transfer 0h and 1w: No significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex task + 33% frequency: KR after every third trial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex task + 100% frequency: KR after every trial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 practice session of 6 blocks of 12 trials (total 72 trials)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Throw with beanbag at target on the floor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The outcome for accuracy is based on the score varied from 0-100 based on zones around the target (higher is better). The variability for accuracy is based on the standard deviation about the mean score (lower is better).
<table>
<thead>
<tr>
<th>Study</th>
<th>TDC (n =)</th>
<th>Focus</th>
<th>Outcome</th>
<th>ANOVA analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wulfs et al., 2010</strong></td>
<td>48</td>
<td>Internal focus: feedback after every third trial with focus on the body (n = 12)</td>
<td>1 practice session of 6 blocks of 5 trials (total 30 trials)</td>
<td>Accuracy: No significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100% frequency with internal focus: feedback after every trial with focus on the body (n = 12)</td>
<td></td>
<td>Quality of movement: No significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>33% frequency with internal focus: feedback after every trial with focus on the body (n = 12)</td>
<td></td>
<td>Quality of movement: No significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100% frequency with external focus: feedback after every trial with focus on ball, target or shoes (n = 12)</td>
<td></td>
<td>Transfer: No significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>33% frequency with external focus: feedback after every trial with focus on ball, target or shoes (n = 12)</td>
<td></td>
<td>Transfer: No significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100% frequency with external focus: feedback after every trial with focus on ball, target or shoes (n = 12)</td>
<td></td>
<td>Transfer: No significant main effect for frequency</td>
</tr>
<tr>
<td><strong>Zamani &amp; Zarghami, 2015</strong></td>
<td>45</td>
<td>Internal focus: No feedback</td>
<td>1 practice session of 6 blocks of 10 trials (total 60 trials)</td>
<td>Accuracy: Significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Post-hoc testing: 100% &gt; 50% **, 100% &gt; 0% ** and 50% &gt; 0% **</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retention: Significant main effect for frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Post-hoc testing: 50% &gt; 100 **, 50% &gt; 0% ** and 100% &gt; 0% **</td>
</tr>
</tbody>
</table>

**Notes:**
- RCT: Randomized Controlled Trial
- TD C: Therapeutic Exercise Center
- Zaman i & Zarghami, 2015: Zaman i and Zarghami, 2015
- Pre-test: During acquisition
- Post-test: Retention: 24h
- ANOVA analysis with post-hoc Tukey HSD tests:
- Acqu i sition: Significant main effect for frequency
- Retent ion: Significant main effect for frequency
- Post-hoc testing: 100% > 50% **, 100% > 0% ** and 50% > 0% **
<table>
<thead>
<tr>
<th>Study</th>
<th>Group 1 (n)</th>
<th>Group 2 (n)</th>
<th>Design</th>
<th>Procedure</th>
<th>Outcome</th>
<th>Pre-test</th>
<th>Retention</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zamaniet al., 2015</td>
<td>TDC: n = 21</td>
<td>Children with ASD: n = 21</td>
<td>RCT</td>
<td>6-8th Inexperienced</td>
<td>1 practice session of 6 blocks of 10 trials</td>
<td>Accuracy: score varied from 0-100 based on zones around the target (higher is better)</td>
<td>During acquisition: Retention: 24h</td>
<td>ANOVA analysis with Tukey-Kramer tests</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TDC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acquisition:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Significant main effect for frequency Post-hoc testing: 100% &gt; 50%<strong>, 100% &gt; 0%</strong> and 50% &gt; 0%**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retention:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Significant main effect for frequency Post-hoc testing: 50% &gt; 100%<strong>, 50% &gt; 0%</strong> and 100% &gt; 0%**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Children with ASD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acquisition:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Significant main effect for frequency Post-hoc testing: 100% &gt; 50%<strong>, 100% &gt; 0%</strong> and 50% &gt; 0%**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retention:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Significant main effect for frequency Post-hoc testing: 50% &gt; 100%<strong>, 50% &gt; 0%</strong> and 50% &gt; 0%**</td>
</tr>
<tr>
<td>Chivascoisky et al., 2008</td>
<td>TDC: n = 26</td>
<td>100% (0.5)</td>
<td>Inexperienced</td>
<td>RCT</td>
<td>Self-controlled: KR on request</td>
<td>Throwing with beanbag at target on the floor</td>
<td>1 practice session of 6 blocks of 10 trials</td>
<td>Accuracy: score varied from 0-100 based on zones around the target (higher is better)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acquisition:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No significant main effect for timing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retention:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Significant main effect for timing 50% &gt; yoked*</td>
</tr>
</tbody>
</table>

** = p < 0.05; ** = p < 0.01; *** = p < 0.001
<p>| Hemayattalab et al., 2013 | Children with CP, GMFCS 1-3 (n = 24) | 11.6 ([1.5]) | Inexperienced | RCT | • Self-controlled: KR on request (n = 10) | • Yoked: KR whenever counterpart requested feedback (n = 10) | Throw with beanbag at target on the floor' | 1 practice session of 10 blocks of 8 trials [total 80 trials] | Outcome | Accuracy: score varied from 0-100 based on zones around the target [higher is better] | Pre-test | During acquisition | Post-test: • No significant main effect for timing (p = 0.473) | Retention: • Significant main effect for timing (p = 0.003) SC &gt; yoked Transfer: • Significant main effect for timing (p = 0.018) SC &gt; yoked | MANOVA analysis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hemayattalab et al., 2014 | Children with CP, GMFCS 1 (n = 22) | 12.26 ([3.11]) | Inexperienced | RCT | • Self-controlled: KR on request in maximum half of the trials (n = 8) | • Instructor-controlled: KR when instructor wanted in half of the trials (n = 7) | • Control: no KR (n = 7) | Dart throwing | 5 practice session of 4 blocks of 5 trials, period not reported [total 100 trials] | Outcome | Accuracy: score varied from 0-100 based on zones around the target [higher is better] | During acquisition | | ANOVA analysis with post-hoc Tukey HSD tests | • Post-test: No significant main effect for timing | • Retention: Significant main effect for timing Post-hoc testing: SC = IC*, SC &gt; control (p = 0.014) and IC = control* Transfer: Significant main effect for timing Post-hoc testing: SC = IC*, SC &gt; control* and IC = control (p &gt; 0.05) | |</p>
<table>
<thead>
<tr>
<th>Form</th>
<th>2012</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD C (n = 40)</td>
<td>10.4 (1.0)</td>
<td>Inexperienced</td>
</tr>
</tbody>
</table>

**RCT**

- **Reduced frequency with faded frequency** from 100% to 0% (n = 10)
- **100% frequency**: KR after every trial (n = 10)
- **Self-controlled**: KR on request in 3 out of every 10 trials (n = 10)
- **Bandwidth**: feedback when score is less than 50 out of 100 points (n = 10)

**Throw with beanbag at target on the floor**

- 1 practice session of 6 blocks of 10 trials (total 60 trials)

**Outcome**

- Accuracy: score varied from 0-100 based on zones around the target (higher is better)

**Analysis with post hoc Tukey**

- Significant main effect for KR feedback

---

**Zaman et al., 2015**

- Children with CD and MID, IQ range 50-70
- (n = 24)

**RCT**

- **Self-controlled + 75% frequency**: KR on request in 75% of the trials (n = 6)
- **Self-controlled + 50% frequency**: KR on request in half of the trials (n = 6)
- **Instructor-controlled + 75% frequency**: KR when instructor worked in 75% of the trials (n = 6)
- **Instructor-controlled + 50% frequency**: KR when instructor worked in 50% of the trials (n = 6)

**Throw with tennis ball at target**

- 8 practice sessions of 6 blocks of 10 trials, period not reported (total 240 trials)

**Outcome**

- Accuracy: score varied from 0-100 based on zones around the target (higher is better)

**Analysis with post hoc Tukey**

- Significant main effect for KR feedback

---

**Sahar et al., 2017**

- Reduced frequency of 100% to 0% (n = 10)
- Timing unclear

**RCT**

- **Self-controlled**: KR on request in 50% of the trials (n = 6)
- **Instructor-controlled**: KR when instructor wanted in 50% of the trials (n = 6)

**Throw with ball with target on the floor**

- 5 practice sessions of 6 blocks of 10 trials (total 300 trials)

**Outcome**

- Accuracy: score varied from 0-100 based on zones around the target (higher is better)

**Analysis with post hoc Tukey**

- Significant main effect for KR feedback

---

**Analysis**

- Treatment: timing unclear

---

**Note:**

- This analysis was conducted with post hoc Tukey.

---

**Analysis**

- Treatment: KR feedback

---

**Analysis**

- Treatment: KR feedback

---

**Note:**

- This analysis was conducted with post hoc Tukey.

---

**Analysis**

- Treatment: timing unclear

---

**Analysis**

- Treatment: KR feedback

---

**Note:**

- This analysis was conducted with post hoc Tukey.

---

**Analysis**

- Treatment: timing unclear

---

**Analysis**

- Treatment: KR feedback

---

**Note:**

- This analysis was conducted with post hoc Tukey.

---

**Analysis**

- Treatment: timing unclear

---

**Analysis**

- Treatment: KR feedback

---

**Note:**

- This analysis was conducted with post hoc Tukey.
<table>
<thead>
<tr>
<th>RCT</th>
<th>Instruction with visual analogy (n = 12)</th>
<th>Instruction with verbal analogy (n = 12)</th>
<th>Explicit instruction: with internal focus on arm or hand (n = 12)</th>
<th>Control: no specific instruction (n = 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td>Accuracy: score varied from 0-5 based on zones around the target</td>
<td>Higher is better</td>
<td>Covariates</td>
<td>Age</td>
</tr>
<tr>
<td>During acquisition</td>
<td>Retention: Transfer: 24h</td>
<td>Pearson correlation</td>
<td>Covariables were not correlated to accuracy score</td>
<td>ANOVA analysis with post-hoc tests with Bonferroni correction</td>
</tr>
<tr>
<td>Retention:</td>
<td>Significant main effect for group</td>
<td>Post-hoc testing: visual analogy &gt; verbal analogy** 95% CI = [1.57, 9.77], visual analogy &gt; explicit** 95% CI = [1.07, 9.27], verbal analogy &gt; control** 95% CI = [5.65, 13.85], verbal analogy &gt; control* 95% CI = [0.02, 8.18], explicit &gt; control** 95% CI = [0.48, 8.68] and verbal analogy = explicit [p = 0.74] 95% CI = [-4.63, 6.60]</td>
<td>Transfer: Significant main effect for group</td>
<td>Post-hoc testing: visual analogy &gt; verbal analogy** 95% CI = [2.50, 9.00], visual analogy &gt; explicit** 95% CI = [1.41, 7.92], verbal analogy &gt; control** 95% CI = [5.91, 12.42], verbal analogy &gt; control* 95% CI = [0.16, 6.67], explicit &gt; control** 95% CI = [1.25, 7.75] and verbal analogy = explicit [p = 0.39] 95% CI = [-4.34, 2.17]</td>
</tr>
</tbody>
</table>
controlled; IC = instructor-controlled; CI = confidence interval; IQ = Intelligence Quotient; WMI = Working Memory Index; SBAR = Social Responsiveness Scale; post-hoc test was used; * p < 0.05; ** p ≤ 0.001.
Table 2. Best-evidence synthesis of instructions or feedback applied with a specific frequency, timing or form

<table>
<thead>
<tr>
<th>Parameter studied</th>
<th>Comparison</th>
<th>Author</th>
<th>Task (Population)</th>
<th>Evidence synthesis per study</th>
<th>Evidence synthesis summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Evidence synthesis per study</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Acquisition</td>
<td>Retention</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>During</td>
<td>Post</td>
</tr>
<tr>
<td>Frequency</td>
<td>Reduced fixed vs continuous</td>
<td>Sidaway et al. 2012</td>
<td>Throw with bean bag (TDC)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zaman &amp; Zarghami 2014</td>
<td>Throw with bean bag (TDC)</td>
<td>C</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zaman et al. 2015</td>
<td>Throw with bean bag (TDC)</td>
<td>C</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wulf et al. 2010</td>
<td>Throw with soccer ball (TDC)</td>
<td>NA</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>De Oliveira et al. 2009</td>
<td>Throw with bocha ball (TDC)</td>
<td>NS</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hemayatlab &amp; Rostami 2010</td>
<td>Dart throwing (Children with CP)</td>
<td>NA</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gillespie 2003</td>
<td>Golf putting (Children with MID)</td>
<td>C</td>
<td>NA</td>
</tr>
<tr>
<td>Frequency</td>
<td>Reduced faded vs continuous</td>
<td>Saiz et al. 2012</td>
<td>Throw with bean bag (TDC)</td>
<td>NS</td>
<td>NA</td>
</tr>
</tbody>
</table>

* CC-BY 4.0 International license

It is made available under a CC-BY 4.0 International license (which was not certified by peer review) and is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
<table>
<thead>
<tr>
<th>Timing</th>
<th>Self-controlled vs instructor-controlled (equal frequency in both groups)</th>
<th>Chiviacovsky et al. 2008</th>
<th>Throw with beanbag [TDC]</th>
<th>NS</th>
<th>NA</th>
<th>24h</th>
<th>SC</th>
<th>NA</th>
<th>NA</th>
<th>X</th>
<th>-</th>
<th>**</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hemayattalab et al. 2013</td>
<td>Throw with beanbag [Children with CP]</td>
<td>NS</td>
<td>NA</td>
<td>24h</td>
<td>SC</td>
<td>24h</td>
<td>SC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zamani et al. 2015</td>
<td>Throw with tennis ball [Children with DCD and MID]</td>
<td>NA</td>
<td>NS</td>
<td>NR</td>
<td>SC</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hemayattalab et al. 2014</td>
<td>Dart throwing [Children with CP]</td>
<td>SC</td>
<td>NA</td>
<td>24h</td>
<td>NS</td>
<td>24h</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Self-controlled vs 100% feedback</td>
<td>Sabzi et al. 2012</td>
<td>Throw with beanbag [TDC]</td>
<td>NS</td>
<td>NA</td>
<td>24h</td>
<td>C</td>
<td>NA</td>
<td>NA</td>
<td>-</td>
<td>NA</td>
<td>*</td>
<td>C</td>
</tr>
<tr>
<td>Form</td>
<td>Visual analogy vs verbal analogy</td>
<td>Tse &amp; Masters 2019</td>
<td>Basketball free throw [Children with ASD and MID]</td>
<td>NS</td>
<td>NA</td>
<td>24h</td>
<td>VisA</td>
<td>24h</td>
<td>VisA</td>
<td>-</td>
<td>NA</td>
<td>*</td>
<td>VisA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variability</td>
<td>Reduced fixed vs continuous</td>
<td>Sidaway et al. 2012</td>
<td>Throw with beanbag [TDC]</td>
<td>NS</td>
<td>NS</td>
<td>1w</td>
<td>NS</td>
<td>1w</td>
<td>NS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>De Oliveira et al. 2009</td>
<td>Throw with bocha ball [TDC]</td>
<td>NS</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Oh</td>
<td>NS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Quality of movement</td>
<td>Reduced fixed vs continuous</td>
<td>Wulf et al. 2010</td>
<td>Throw with soccer ball [TDC]</td>
<td>NA</td>
<td>NS</td>
<td>24h</td>
<td>C</td>
<td>Oh</td>
<td>24h</td>
<td>C</td>
<td>C</td>
<td>NA</td>
<td>-</td>
</tr>
</tbody>
</table>

Legend: h = hour(s); w = week(s); TDC = typically developing children; ASD = autistic spectrum disorder; CP = cerebral palsy; MID = mild intellectual disabilities; DCD = developmental coordination disorder; NA = not applicable; NR = not reported; C = significant, favouring continuous frequency; R = significant, favouring reduced frequency; SC = significant, favouring self-controlled feedback; VisA = significant, favouring the visual analogy. Consistency was defined as 75% of the studies assessing the same comparison showing results in the same direction. Strength of the evidence according to the guidelines of van Tulder et al.:

*** = Strong – consistent findings among multiple high quality RCTs.
** = Moderate – consistent findings among multiple low quality RCTs and/or CCTs and/or one high quality RCT

* = Limited – one low quality RCT and/or CCT

X = conflicting – inconsistent findings among multiple RCTs and/or CCTs.

- = no evidence from trials – no RCTs or CCTs