

Vebjørn Brauten Urdal

Contextual Interference in Soccer

A systematic review and a Meta-Analytic study

Master thesis in Sport Sciences

Department of Coaching and Psychology
Norwegian School of Sport Sciences, 2016

Abstract

Soccer is one of the biggest sports in the world, and the level of expertise at the top level is high. The development of this level of expertise takes time, but it also depends on the quality of practice. The purpose of this research thesis was to systematically review how practice schedules in Soccer affects learning, differing in degrees of contextual interference (CI). A secondary method was included, a meta-analysis, to determine the magnitude of the relationship, between the degree of CI and learning. Meta analysis calculate a total effect of studies, and therefore is a better estimate than a single study. The CI effect, first noted by Battig (1966), states that a high degree of CI facilitates the application of learning. A suggestion, is a division along two continuums, describing the degree of contextual interference and variability in practice schedules (Williams & Hodges, 2005). Both the degree of CI and variability is affected by the demands in the practice schedule, where higher degree is more demandable (Shea & Morgan, 1979). CI is used as the common term for both contextual interference and variability in the remainder of this section.

This thesis included 17 studies, where 12 reported significant differences in learning, almost exclusively in the direction of higher degree. The magnitude of relationship was found with a large effect in general across effects, supporting the original notion of CI effect. This general effect was homogeneous, as were the large effect of the outcome decision making. These results are robust, and have clear practical implication. Across moderator variables not found homogeneous in the meta-analysis, the results support high degree of CI, including; age, outcomes, stability of environment testing procedure, testing time point. Some studies included three practice groups, but as the meta-analysis only can compare two practice groups, the group in the middle were both low and high depending of the comparison. Though, were the middle group (moderate degree of CI) compared qualitatively with the others, showing promising learning effects also compared to higher degree, when in a stable environment of treatment. Future research should continue to increase the depth, as this thesis and earlier research on a broader domain (Brady, 2004) both support high degree of CI.

Table of contents

Abstract	3
Table of contents.....	4
Preface.....	6
1. Introduction.....	7
1.1 Soccer – context and expertise.....	9
1.1.1 The nature of Soccer.....	9
1.1.2 Expertise in Soccer.....	10
1.2 Developing expertise in soccer.....	13
1.3 Theoretical learning approaches	15
1.3.1 Traditional approach	15
1.3.2 Teaching Games for Understanding approach	16
1.4 Contextual interference in soccer.....	17
1.4.1 Explaining the contextual interference effect	18
1.4.2 Origin and effects of contextual interference	19
1.4.3 Stability of environment.....	20
1.5 Aim of study	21
2. Method.....	23
2.1 Search strategies.....	24
2.2 Inclusion/ exclusion criteria	25
2.3 Data extraction	27
2.4 Analysis	27
2.4.1 Systematic review.....	27
2.4.2 Meta-analysis.....	28
3. Results.....	34
3.1 General results.....	34
3.1.1 Studies with three treatment groups	37
3.2 Main results - Meta analysis.....	38
3.2.1 Correlates below threshold	43
4. Discussion.....	44
4.1 Summary of learning effects	44

4.2	Learning effects	45
4.2.1	Overall	45
4.2.2	Stability of environment	46
4.2.3	Outcomes	47
4.2.4	Testing time point	50
4.2.5	Age	51
4.2.6	Three groups of comparison	52
4.2.7	Lack of learning effects	53
4.3	Strengths and weaknesses	54
4.3.1	Search and selection process	54
4.3.2	Analysis/ results	59
5.	Conclusion	65
6	References	66
7.	List of Figures	93
8.	List of Tables	94
9.	Abbreviations	95

Preface

The last five years I have been studying at the Norwegian school of sport science, and hopefully ending with a master degree in sport science. No matter grade, whether singing songs of ABBA or DDE, tomorrow will be a time of relief and feeling wistful. Whatever the future brings, I will look back at the past years with good memories.

The Norwegian school is for me first and foremost a place of people. In these years I have friended, befriended, argued, gotten rejected, but mostly enjoyed socializing when not studying. The academics has been a interestingly, and absolutely tiresome acquaintance, that surpassed previous years when writing the master degree.

As I have been travelling between towns, and at times countries, I chose to do a systematical review, with the aim of being independent for a year. The conduction of this, especially the meta analysis part was interesting, but hard, and I don't think I would have finished it alone. So much for independency.

The biggest thanks goes to my supervisor Henrik Gustafsson. An incredible positive person and inspiration. But also other people have contributed tremendously. My uncle, John Arne Skolbekken has contributed with notes, advices and lunches in the cantina of NTNU. Bård Erlend Solstad, soon to be Doctor Solstad, and the experienced statistician Ingar Morten Holme, have helped with the quantitative measures.

As for social support, my roommate Joakim Hesjevik, having no other choice than to socialize irrespectively the time of day. At last I have to thank the most important people in my life, respectively my my mom, dad, brother and sister, as well as to my friend and coach Sarmed Saify. In this period, especially the last six months they have aided financially and provided shelter, as well as being there for me.

For those I forgot, give me a heads up. I'll buy you one.

Oslo, may 2016

Vebjørn Brauten Urdal

1. Introduction

In sport, like in many domains, expert performance is of the highest order, and is the main goal. The spectators get amused by the ones demonstrating their skills. In the background are the facilitators, not just the coaches, or the physio and the rest of the staff who are involved directly, but also the owners' board and administration at different levels.

Soccer is one of the most popular sports worldwide. In total 1,4 billion people of the world's population have an interest in the sport of soccer (Giulianotti & Robertson, 2004), and 270 million are actively involved on a regular basis. That is four percent of the total population in the world. This amount has increased by 10 percent, in the period from 2000 to 2006 (FIFA, 2007). The biggest, most spectacular, event at club level is the Champions League where 32 European clubs qualify and compete for the championship. A mere 0,0001 percent of the active participants in the world of soccer contribute by playing in the group phase, and naturally even fewer in the knockout rounds when the level increase even more. So what does it take to develop that level of expertise?

The purpose of this research was to systematically review studies that compared different practice schedules, to determine their impact on expertise development in soccer. Practice schedules in the studies departed from each other, by their amount of contextual interference and variety. As a secondary method, a meta-analysis was conducted, to determine a more meaningful review of the studies. Meta analysis, calculating the total effect of the studies, is a better estimate of effect than single studies.

With the number of people interested in soccer, this part of research on how to develop expertise is naturally a respected area in sports science (Williams & Reilly, 2000). Unfortunately, research in soccer has for a long time been out of favour with researchers because of the challenge of measuring and assessing team sports (Reilly & Gilbourne, 2003). In recent times, however, the research on elite performance has increased

significantly increased in recent time, for a review on the practice, instruction and skill acquisition in soccer see Williams & Hodges (2005).

Developing expertise, is not only of interest for the clubs who participates in the Champions League, as the biggest clubs got money to buy this level of expertise, but also for the smaller clubs. The biggest clubs participate in the Champions League who is a financial gold mine, making the rich clubs even richer. The tournament got the biggest prize money of all soccer competitions, 1,2 billion Euros, and TV-revenue money, 480 million Euros (UEFA, 2015). This does not make the biggest clubs neglect development, but rather give them two options. Some smaller clubs have the goal of developing expertise, with the motive of selling players as an income for their club, or they want to win smaller tournaments, but most clubs and individuals plainly has the goal of human development, or want to contribute to soccer's development as a sport.

Regardless of the desire for expertise you have the two already mentioned options. The buying option of expertise is beyond the purpose of this thesis, however, as is the subject of identifying talented players, although the possibility of identifying such players for development purposes (Williams & Reilly, 2000), whose potential for expertise may be developed through the right practice schedule.

Within the expertise development research, the nature versus nurture debate has been ongoing for a long time (Ericsson, 1996; Ericsson & Lehmann, 1996). This discussion and other developmental areas will be discussed, but first it is necessary to elaborate what differentiates soccer from other sports. In the lack of studies in soccer appropriate for selection with this thesis scientifically approach, design, and purpose, the need to include studies from other sports are inevitable. Therefore, clear guidelines on which sports that are appropriate comparisons will be discussed and decided, based on commonalities in the available literature. In relation to this, the need for elaborating which skills that differentiates between levels of expertise is included.

1.1 Soccer – context and expertise

1.1.1 The nature of Soccer

In Soccer, individuals get teammates. Team sports, as the term literally demonstrates, is used as the common term for sports where you compete together with others for a shared goal. Unlike, individual sports, where you compete on your own. This is the first distinction. Furthermore, soccer can be separated from other team sports by looking at its complexity in using the same space as the opponents. In soccer, as it is an invasion game, to invade your opponent's areas with the purpose of scoring goals is a specific goal (Rongland, 2008). The continuous process before and during the execution of an action, including the need to anticipate the environment, where you are depending on your teammates' helping efforts, and the opponents are trying to stop you, as you make decisions. This differentiates soccer from net/wall games with sports like tennis and volleyball, and the term for this kind of sports is invasion sports or territorial games (Griffin & Butler, 2005).

Other types of sports, who are different from invasion sports, have other common terms and objectives. Net/wall games, already mentioned, have as their goal to strike an object over the net/wall into the other team's separated area, in a way that makes it unreturnable (Mitchell & Oslin, 1999). In striking/ fielding games the opponents play in the same areas, but there are breaks between plays allowing the players to recuperate. Such games are even less complex than the other games mentioned. Each player/team has their own area where there is no intersection between the opponents (Griffin & Butler, 2005).

Invasion games include sports like basketball, handball, Lacrosse, netball, rugby, speedball, ultimate Frisbee, water polo, different kinds of hockey including field, floor and ice hockey, and of course soccer, which is the research objective in this thesis. Transfer effects between sports, including between invasion sports, holds that experience of previous practice in a domain can enable successful performances in other domains (Duncan, 1953). There are not extensive amounts of research in this domain. For instance, there is a lack of research investigating whether decision making skills are transferable across sports (Causer & Ford, 2014). The research done in this domain,

despite a limited amount of studies, support the transfer across invasion sports hypothesis.

Smeeton, Ward & Williams (2004) found transfer effects between the invasion sports soccer and field hockey. The skilled players' ability to recognize structures in both sports were quicker than the less skilled participants. The same transfer effect was not found for volleyball players, or when skilled soccer and field hockey players tried recognizing volleyball structures. Pattern recognition is associated with the ability to make decisions in sports (Causser & Ford, 2014). Causser & Ford (2014) have found similar positive transfer effects, between soccer and other invasion sports, when testing decision making. Sports named were basketball, hockey, and rugby, but the transfer effect was not found for the other categories of sports, including tennis, golf and athletics. To sum, the ability to transfer decision making across sports, more specifically strategies and perceptual information, across invasion games was found, but only to a limited extent. The transfer was not found for net/ wall games like volleyball and tennis, the target game golf and for the individual sports in the track and field domain. This supports the already mentioned differences between the categories, pointing in the direction of the inclusion of other invasion sports than soccer for the purpose of this study.

1.1.2 Expertise in Soccer

Knapp (1977) has argued in favour of the importance of decision making in soccer, saying that a player is almost useless without the ability to make the execution at the right time, although the player is able to execute the skill technically. The ability to make appropriate decisions has been tested in soccer between teams, with different levels of expertise. Decision making can be defined as the ability to both perceive and interpret information in the playing environment, and then select the appropriate decision (Baker, Cote & Abernethy, 2010).

Fully matured players, aged 18-23, were tested in tactical variables involving the ability to make decisions. The team with the higher level of expertise, the Netherlands, outperformed the lower level, the Indonesian, at knowledge about ball actions, other actions and positioning and deciding, tested by means of a questionnaire questionnaire (Kannekens, Elferink-Gemser & Visscher, 2009). The use of testing procedures, like

questionnaires for instance, outside the sport environment has methodical implications to be aware of. This matter will be discussed in more detail later. The same study of Kannekens and colleagues (2009) found differences between the first-eleven players compared to other players in the Indonesian team, where positioning and deciding had a positive association with playing time (Kannekens et al., 2009).

Ali (2011) has examined both the use of different testing procedures, and more interestingly in this section, the differences between levels of expertise. It has to be noted that college students are used in this research, though. This is not necessarily a problem, but a possible implication to be aware of, as it is not necessarily reflecting a high level of expertise.

Traditionally in some countries, like in the USA where the study of Ali (2011) was conducted, sport activities are mainly organized within the education system and the studied athletes are students. Another option used, in countries like Norway, is the organization of sports in clubs based geographically where people live. Both these options are mainly a result of convenience, but at some stage, often in the latter part of adolescence, both clubs and schools start recruiting people by their level of expertise.

A third option, often includes subjects with the highest level of expertise, and may therefore be the best subjects when researching differences between levels (Ali, Williams, Hulse, Strudwick, Reddin, Howarth, Eldred, Hirst & McGregor, 2007), as done by Kannekens and colleagues (2009). This procedure of development, frequently used in the best soccer countries, recruit people into the academies organized by the biggest clubs from an earlier age. It is not necessarily the best option for development, but it is the best option for recruiting a higher level of expertise into research when the highest level, like Champions league players, often are out of reach. Because the highest level, like Champions league players, are often out of reach. An advantage of the research done by Ali (2011) is that subjects' age, was above maturation age of 18 years old, unlike most research comparing different levels of expertise (Deprez, Franssen, Lenoir, Phillipaerts & Vaeyens, 2015; Figueiredo, Goncalves, Coelho E Silva & Malina, 2009; Huijgen, Elferink-Gemser, Lemmink & Visscher, 2014; Le Gall, Carling, Williams & Reilly, 2010; Reilly, Williams, Nevill & Franks, 2000).

All the studies included in the review of Ali (2011) tested performance outcomes in soccer. When testing for perception and anticipation differences, differences were found between levels of decision making (McMorris & Graydon, 1996) and memory according players' positions (McMorris & Graydon, 1997). This supports the already stated notion of the importance of tactical variables and the development of decision making, but Ali (2011) has also shown differences at execution variables. With regard to technical variable differences, between elite and other level of expertise, there has been found inequalities in; shooting (Ali et al., 2007), passing (Ali et al, 2007; Ali, Foskett & Gant, 2008) and dribbling (Rösch, Hodgson, Peterson, Graf-Baumann, Junge, Chomiak & Dvorak, 2000).

Maturation will impact the test results before the age of 18, and is the reason behind the choice of literature presented here, about expertise differences. In a study done on soccer players, 39 percent of the players aged 17 years old, at the highest level, had completed their biological maturation. The remaining 61 %, was, or were on schedule, of finishing maturation, at the age of 18 years old (Malina, Pena Reyes, Figueiredo, Coelho E Silva, Horta, Miller & Morate, 2010).

Relative age effect (RAE) is the implication in the years before completed maturation, also in soccer (Helsen, Winckel & Williams, 2005). RAE is defined as "Relative age refers to the differences in ages between children in the same age group that result from their different birthdates throughout the year" (Barnsley, Thompson & Legault, 1992, s78). Since children born in the same year can be close to twelve months apart in chronological age, these age differences may also be reflected in their cognitive, physical and emotional maturity (Musch & Hay, 1999). It has been demonstrated that athletes born later get excluded, an especially in the years of growth spurt (Figueiredo et al, 2009; Figueiredo & Coelho e Silva, 2010; Malina, Pena Reyes, Eisenmann, Horta, Rodrigues & Miller, 2000).

Now, after operationalizing the background for including and defining sports, as well as giving a backdrop of skills differences between levels in soccer, time has come to give an overview on how to develop these skills. A learned skill is an ability to execute a given action, with the greatest possible probability, often with as little consumption of energy and time as possible (Knapp, 1977).

1.2 Developing expertise in soccer

When looking at skilful players' people often refer to elite players like they got a gift for the game, implying that they have unique abilities which make them destined for excellent achievements within the sport. The classic debate between the influence of genes (nature) and environmental factors (nurture) on developing human behaviour is historical and ongoing, but the extreme versions have been increasingly greyed out (Davids & Baker, 2007). At this point there is little support for a particular influence alone, but there is evidence rejecting the idea that individuals can have a biological disposition from single genes that lead to superior performance in a specific domain (Davids & Baker, 2007). Most researchers are in agreement on the importance of the role of deliberate practice for developing expert levels of performance, but they still discuss if it sufficient alone (Baker, 2007).

Deliberate practice is typically defined as a structured activity with the specified goal of developing domain specific skills, as opposed to play activity (Ericsson, Krampe & Tesch-Roemer, 1993). The deliberate practice term was presented as the Expert Performance Approach by Ericsson and Smith (1991), later reframed to the deliberate practice framework (DPF) by Ericsson and colleagues (1993). DPF originated from experiments concerning developing expertise in violin playing, and proposed that the amount of practice with good quality is the most discriminating factor for the development of expertise (Ericsson et al., 1993), and not necessarily just quantity alone.

The same notion of differences between levels of expertise and amount of deliberate practice has thereafter been confirmed in other domains, which have facilitated the framework (Ericsson & Lehman, 1996). Recent research by Haugaasen, Toering & Jordet (2014) in soccer, support that the amount of total practice alone is not differentiating between levels, when no significant differences were found between professional players and non-professionals from age of 6 to 19 years old. Differences in accumulated hours were found in some specified forms, namely play activities and coach-led practice at the age of 6 to 8 years old.

The lack of existing literature supporting a notion that the amount of practice alone is of significant importance, neglecting the quality, makes it necessary to look into the quality of the practice. How should the practice be structured relative to facilitating

effective learning? This is crucial for coaches and often their responsibility, but can also be seen useful information for a range of different operators in soccer.

Ward, Hodges, Williams & Starkes (2004) have pointed to the lack of knowledge about specific practice activities, and their impact on the attainment of excellence. It has been noted that especially how and what to practice are questions in need of further exploration (Janelle & Hillman, 2003). Although lack of research, there are some guidelines based on empirical evidence in motor research, on how to structure practice schedules. This framework is the theoretical background in this thesis, pointing out that constant, specific and blocked practice schedule is better at facilitating performance in the present moment, while random and variable practice is a more effective approach for learning and retaining skills (Williams & Hodges, 2005). Hopefully, when looking at the framework in the specific form of invasion games, clearer implications can facilitate the future of coaching practice.

Coaching practice nowadays is often determined subjectively, referred to as intuition by practitioners. This is the traditional backdrop for how to practice, rather than depending on empirical research (Williams & Hodges, 2005). If the practice schedules used in the coaching practice were the same as the empirical evidence, a subjective approach based on intuition would not be a problem. This is not the case, however. Contrary, a gap between practice and empirical evidence is reported. In a study of 25 coaches during 75 practice sessions, most of the time was spent in activities not replicating the demands of the competition, practice which does not facilitate learning in the same extent as practice in a competitive environment (Ford, Yates & Williams, 2010).

With regard to the difference of performance and learning, it is necessary to operationalize both terms. Performance is used when referring to observed behaviour. In soccer, this is typically the behaviour demonstrated in practice or at a match. As a function of practice, learning is a set of processes when experiencing at practice that lead to relative permanent changes in the ability to do an action or movement (Schmidt & Lee, 2011; Stratton, Reilly, Richardson & Williams, 2004). Learning will probably make changes in the nervous system, but cannot be observed directly at a certain point in time, rather it has to be evaluated by assessing changes in performance at various time points (Stratton et al., 2004).

To simplify the difference in soccer, even though a player performs well at practice, that is not necessarily tantamount with indications of learning. Short-term performance is not the same as long-term learning (Stratton et al., 2004). This should be an implication for the coach to be aware of when structuring practice schedules, where the objective is facilitating learning.

1.3 Theoretical learning approaches

1.3.1 Traditional approach

An approach used when developing athletes has been the traditional approach (TA), also known as the technical model or the skill based approach (Williams & Hodges, 2005). In the domain of sport, this approach is categorized by a very specific focus on learning the execution of skills and techniques. Often this is done in a highly structured session with low amounts of context, where isolated exercises are dominant. Typically warm up, skill based practice and a concluding game at the end (Allison & Thorpe; Wright, McNeil, Fry, Wang, 2005). The basic assumption of TA is that athletes, before they get exposed to the game, should get the opportunity to learn and refine the action outside the game context (Williams & Hodges, 2005). An example is the cognitivist viewpoint, in which it is assumed that when the functionality increases, this reflects the traditional approach learning schedule (Tallir, Lenoir, Valecke & Musch, 2007). Cognitivists view the core of human behaviour to be mental capacities of the brain or mind. Criticism of this position has been that the internal view on mental processes comes at the expense of the social surroundings in the environment (Arponen, 2013).

Interestingly, TA makes the assumption that, if an athlete can perform the technical execution of a skill, he will also be able to perform the action appropriately in the game context (Chatzopoulos, Tsormbatzoudiz & Drakou, 2006b). This assumption is at best questionable, as the technical execution is only a part of the total execution in game context, where making suitable decisions on the basis of the environment is an integral part of the overall execution (Grehaigne, Godbout & Bouthier, 2001). When the athlete is developing skills in a context outside the game, the opportunity to learn in which situations the action is fitting are lost (Gray & Sproule, 2011). Another drawback is that the lack of context removes the creativity and opportunity to be actively involved in the learning process (Griffin & Butler, 2005).

1.3.2 Teaching Games for Understanding approach

As an opposing approach to the TA the Teaching games for understanding model (TGfU) was presented. First presented 30 years ago, originally by Bunker & Thorpe (1982), the TGFU model has its origin as a framework for teachers in Great Britain (Thorpe, Bunker & Almond, 1986). The framework was originally made as a practical framework made by practitioners, not as a theoretical teaching model. It can be seen as an idealistic approach, where engagement and interest in the learning process by the learner are focussed (Butler, 2014).

Three main ideas about games is the underlying core in TGFU (Griffin, Brooker & Patton, 2005). Firstly, games can be modified with the goal of elaborating a tactical problem (Griffin et al., 2005). For example, rules in the game (Abrantes, Nunes, Macas, Leite & Sampaio, 2012), the size of the field (Silva, Duarte, Sampaio, Aguiar, Davids, Araujo & Garganta, 2014) and the number of players (Abrantes et al., 2012; Holt, 2006). Secondly, an assumption already elaborated earlier in this thesis, that different games share tactical problems, and therefore a solution is transferable to other types of games. This depends on organizational structures, but can potentially give learning outcomes between sports (Griffin et al., 2005). The third idea, proposes that the game by itself is a natural source of feedback and further development. Through reflection and self-regulating behaviour, the athlete may learn from the game in an authentic context (Griffin et al, 2005). This is at least partly supported, both by studies on self-control (Wulf, Shea & Lewthwaite, 2010), and on self-regulating behaviour researched in soccer at youth ages (Toering, Elfering-Gemser, Jordet & Visscher, 2009).

TGfU has gradually been spread around the world, and similar approaches has arisen like Play practice (Lauder, 2001) and Games Centered Approach (Griffin, Mitchell & Oslin, 1997). As a result of the disperse of TGfU and similar approaches, it has been proposed to form a common term for this direction of learning theory. Games Centered Approach has been proposed as a more including name than TGfU, but the decision on which term that will be the common one across these will first be made in July 2016 (Butler, 2014).

Games Centered Approach (GCA) focuses on increased understanding through game like situations already at the beginning of the practice, as well as a concluding part in

the end (Griffin et al., 1997). Learning skills is not of second interest, it is rather an integral part in the game specific context. The beginning game is accompanied by plenary discussions, lead by the coach, where the focus is self-reflection on how the game went. The self-reflection is supposedly a backdrop for the athletes, when exercises without game play are presented (Griffin et al., 1997). Suggestions, propose a context as similar to the game as possible also in this part, even though it is not the competitive game context (Lauder, 2001). Concluding the practice schedule, as mentioned, the game context reintroduced (Wright et al., 2005).

Play Practice (PP), was introduced by Lauder (2001) with both similarities and inequalities to TGfU. Both approaches emphasize the importance of the learners' understanding of the game. However, where TGfU is not focussing on technical training as long as the pupil does not feel the need, PP proposes that the execution and understanding can be learned simultaneous (Holt, Ward & Wallhead, 2006). Lauder (2001) suggests that as an increase in similarities between training and the actual game occurs, the chance of learning will increase in a similar fashion. This suggestion is supported by Brobst and Ward (2002) in their research of female soccer players.

These approaches, TGfU and the similar ones described above, are ecological dynamic approaches to motor learning (e.g. Gibson & Gibson, 1955). In the learning process, the learners' relationship with the environment is the basis for functional solutions. Highly expertise performers, adapts to the shifting environment constraints given in the competitive game context, when successfully executing an action (Davids, Araujo, Hristovski, Passos & Chow, 2012).

1.4 Contextual interference in soccer

Different practice schedules affect motor skills in different ways. A suggestion is a division of schedules along two continuums, describing the degree of contextual interference (CI) and variability (Williams & Hodges, 2005). The practice schedules, including soccer specific schedules, degree of CI and variability affects both performance, and learning acquisition. The CI continuum ranges from blocked (lower degree of CI) to random practice (higher degree of CI), whereas the variability from constant (lower degree) to variable practice (higher degree) schedules (Williams & Hodges, 2005). For the remainder of this thesis, CI will be the common term for both

contextual interference and variability, when variability is not specified in the current section.

The term CI originates from verbal learning research, and refers to the degree of interference generated under practice when two or more tasks are practiced simultaneously. The degree of CI is affected by the task. Typically, if the task demands one or few skills, the CI is of lower degree. In cases where the task demands several skills, the CI in the practice schedule is of higher degree (Shea & Morgan, 1979). The variability of the practice schedule describes the variations, in the task demands, of the skill or skills used (Stratton et al., 2004).

In this thesis the treatment group with a more randomized or variable practice schedule are in the category of high degree of CI, while groups with a schedule more blocked or constant are in the lower degree of CI category. When studies include three groups of treatment, the group falling in between will, in case of CI and variability, be categorized as moderate degree of CI. As the studies in this review do not use the same term, more specifically do not all use CI or variability terms, an overview of classifications are shown in Table 1. The classification is based on the treatments practice schedules described in the studies (i.e. teaching games for understanding as higher degree of CI, and skill based approach as lower degree).

1.4.1 Explaining the contextual interference effect

Schmidt's (1975) Schema theory has traditionally been used when describing the importance of cognitive processes in motor acquisition. It explains that variability in movement and contextual characteristics are important when learning, because it develops bigger and more resistant behaviour, including motor skill, to deal with the variation of task demands in different situations. Two main theories try to describe the cognitive effect of contextual interference in greater detail.

The action plan reconstructing hypothesis (see, Lee & Magill, 1985), builds on Schmidt's Schema theory, and proposes that random practice provides a short period of forgetting, as a result of interference between tasks, which results in the need of reconstructing between each trial (Lee & Magill, 1985). In a blocked manner there is no

need to reconstruct between trials because the task is the same every time (Schmidt & Lee, 2011).

The elaboration hypothesis (see, Shea & Morgan, 1979 or, Shea & Zimny, 1983), is different, indicating that random practice promotes more comparative analysing between the different skills practiced, providing possibilities for exploration on how information and behaviour interacts. The repetitive nature of blocked practice is proposed to foster less analysis (Schmidt & Lee, 2011; Davids, Williams, Button & Court, 2001).

Future research is needed to further determine which cognitive mechanisms that lead to the CI effect (Broadbent, Causer, Williams & Ford, 2015), although both argues that random practice engages the individual in a strenuous process which facilitates the need of retrieval (Young, Cohen & Husak, 1993). No matter the mechanisms of the CI effect, the phenomenon has been relevant since first identified 50 years ago.

1.4.2 Origin and effects of contextual interference

The CI effect was first reported in 1966, and as mentioned when describing the term, the study investigated practice of verbal knowledge (Battig, 1966). In motor learning Shea and Morgan (1979) were the pioneers in motor learning, which has since then been studied extensively. Basic studies have given powerful conclusions (Brady, 1998; Brady, 2004; Magill & Hall, 1990; Van Rossum, 1990), which give clear implications for practitioners of motor skills (Magill, 1992). The conclusions state that practice with several tasks at hand in a random order provide detrimental effects on performance, compared to blocked practice. In contrast, random practice provides superior retention and transfer effect, in other words learning (Brady, 2008). Though it is notable that there are studies which do not support this effect (Pease & Pupnow, 1983; Wrisberg & Mead, 1981), and applied research has provided weaker results than the basic research (Brady, 1998; Brady, 2004; Magill & Hall, 1990).

In a meta-analysis of the effect on CI, Brady (2004) found an average effect size of .38 using Cohen's *d*, when including both basic and applied research from 61 studies. A significant difference between basic and applied research was reported, .57 and .19 respectively. The effect size for adults were also significantly higher (.50) than for

youth (.10). The small effect for youth is not always present. In another review the effect on learning has been showed to decrease with age, contradicting the differences in age found by Brady (2004). When evaluating effect sizes Cohen (1988) consider .20 as small, .50 as moderate and .80 and above as large. Age is not the only moderating factor reported in the research on CI. General motor experience, gender, personality, skill, task and the degree of CI also seems to affect learning differentially (Brady, 1998; Shapiro & Schmidt, 1982).

1.4.3 Stability of environment

As a plausible explanation for these moderating effects on the CI phenomena, Guadagnoli and Lee (2004) has made a concept which separates along two continuums. Nominal difficulty reflects the degree of difficulty regardless of who is executing or the conditions of the task. This type of difficulty is fixed. Functional difficulty, on the other hand, is dependable on both the skills of the one executing and the conditions of the task. This type of difficulty is dynamic in nature.

In a recent study by Cheong, Lay & Razman (2016) the contextual stability is discussed, and a framework presented. The stability refers to the functional difficulty of the demands in form of the conditions (Cheong et al., 2016).

Athletes exposed to random practice in a stable environment (under practice at stable conditions) can be as stable as athletes exposed to blocked practice. When the environment allows practicing in a self-paced tempo without unstable components, the functional difficulty in form of conditions will be similar for random and blocked practice. When the environment has unstable conditions, however, the functional difficulty increases, and the practice is random based on the environment (Cheong et al., 2016). The same notion of unstable environments should be applied when choosing testing procedure. Outcomes dependable on perceptual variables should use experimental designs, representative of the typical environment of performance, when generalization is the intention (Brunswik, 1956). The stability of environment was a potential moderator in the results, and can facilitate the progress of better understanding the CI effect.

Especially, when studying soccer, the stability of the environment is relevant. All sports use cognitive, perceptual and motor skills in various degrees. The environment in soccer is constantly changing, making the demands of all skills mentioned above very high (Bate, 1996). Other sports, for example target games are, as described earlier (Griffin & Butler, 2005), being exerted in a very stable environment and may react differently to degrees of CI when developing athletes. Therefore, it is necessary to distinguish soccer from other sports not classified as an invasion sport when reviewing.

Likewise, as other sports respond differently on the CI effect, different methodological factors in a single sport may respond differently. The stability in the environment of the practice schedule chosen in the intervention, and the testing procedure pre and post to the intervention, may moderate the results and the implication for practitioners. The potential of awareness of these methodical factors when computing the power of the results will give a more valid understanding.

1.5 Aim of study

As mentioned earlier, and in conjunction with understanding, a lack of literature has been noted by some, both generally and more in depth (Janelle & Hillman, 2003; Ward et al, 2004). Literature in a general manner has been available for some time, and both Magill and Hall (1990) and Van Rossum (1990) shed light on the domain of CI effect early in the nineties through reviews. More recent reviews have been completed to give overview (Barreiros, Figueiredo & Godinho, 2007; Brady, 2008), and the newest in 2011 (Merbah & Meulemans, 2011). Literature in more depth, specifying how the contextual interference affects practice in specific domains are lacking. In relation to this, the title of this thesis is; “Contextual interference in soccer – a systematic review and a meta analytic study”, an in-depth study of the effect of CI in soccer practice.

In this thesis, a systematic review which are described in detail in the section of method, was used with the purpose of obtaining a qualitative overview of how practice with different degree of contextual interference affect learning in soccer. In the lack of studies in soccer, similar sports were included and evaluated. The review shows general descriptive data about key items methodically, and in terms of results. Both studies with and without significant results, are described with key features, to give a meaningful overview of the domain.

To further elaborate the true meaning of the results, the ones found significant was conducted through a measure of effect size, respectively Cohen d (Cohen, 1988). As a combined effect gives truer meaning of effect, the estimates of effect sizes from single studies were combined, conducting a meta-analysis, to determine the total magnitude of all studies. More specially, the strength of the relationship between practice schedule and learning effects. In relation to this a computer program was used (comprehensive meta-analysis (3.0)) which is specialized for meta-analysis.

When analysing the effects quantitatively, the data will be put through comparisons looking for potential moderators affects. As a result of combining the meta-analysis with the systematic review, hopefully the question of, how to structure practice with the purpose of maximizing learning, will be answered in a meaningful way, with a calculation of the power of the relation. In the case of inconclusive evidence, at least the descriptive data can give indications of the needed future research directions. This is the second purpose of systematic reviews.

2. Method

Reviews showing the available research and its implications are fine, but by doing it more systematically it will both facilitate the methodical quality of the thesis, and the understanding. About the quality of reviews, it has been an intention to develop this for some time, "*As the amount of research grew, so did the need for credible ways to integrate their findings*" (Cooper, 2010, s. 2). Reviews seek to identify studies, evaluate them and last, but not least, sum the results. All with the goal of making the results more easily available.

Systematic reviews are similar, except that the process is done in a very strict manner. Done openly and accessible for others, specified beforehand and leaving replication as an option (Higgins & Green, 2011). Additionally, systematic reviews demonstrate where knowledge is lacking (Dixon-Woods et al., 2006). A systematic review on the CI effect will reduce the lack of accessible literature in the present time showing directions and implications, or including identifying themes still in need of research.

Results, even when significant, may prove misleading as basis for generalization (Brady, 2004). Cohen (1988) argues that small differences can be declared significant independent of the actual power or the direction it possesses. A possible solution is to combine results, instead of depend on a single study, giving a better estimate of the actual intervention effect (Murtagh et al., 2007; Roen, Arai, Roberts & Popay, 2006). A method which does just that is the meta-analysis, which can demonstrate the accumulated effect from several studies. As with systematic reviews, the meta-analysis has emerged in recent years since their systematic ways took root in the 1990s, replacing more narrative reviews (Borenstein, Hedges, Higgins & Rothstein, 2009).

Before statistical power, the phenomenon in a population was considered as present or absent, and was missing the specific value for a population parameter (Cohen, 1988). Power analysis exploits the relationship among population effect size, sample size, significance criteria, and statistical power (Cohen, 1992).

Meta-analysis includes measures of statistical power like effect size. Effect size, as a statistical power, is an objective measuring of meaning, when combined with statistical

significance, which results in more precise estimates of an intervention's true effect (Brady, 2004). Meta-analysis is based on quantitative methods to sum effects size, and show strengths of relationships. This separates meta-analysis from systematic reviews. Though, when seen appropriate, systematic reviews can be elaborated in more detail by meta analytic methods (Alderson & Green, 2008).

2.1 Search strategies

An electronic search was done in March 2016, with the purpose of finding studies compatible with the aims of this thesis systematical review. Four databases were chosen; specific databases in sport (SPORTDiscus), medicine (PubMed), in addition to an interdisciplinary database (Web of Science) and the multi resources database Ovid (Journals@Ovid full text, NIHs Journals@Ovid, AMED, Embase, Epub Ahead of print, Ovid MEDLINE(R), MEDLINE and PsycINFO). Google Scholar was not used, as it does not give additional findings and is very time consuming. (Higgins et al., 2011). The keyword combinations are shown in Figure 1, and were chosen after critical discussion with my supervisor. An extended search was performed by reviewing the chosen studies' references, but no additional studies matching the inclusion criteria were found.

For a more detailed overview of the search, see Figure 1.

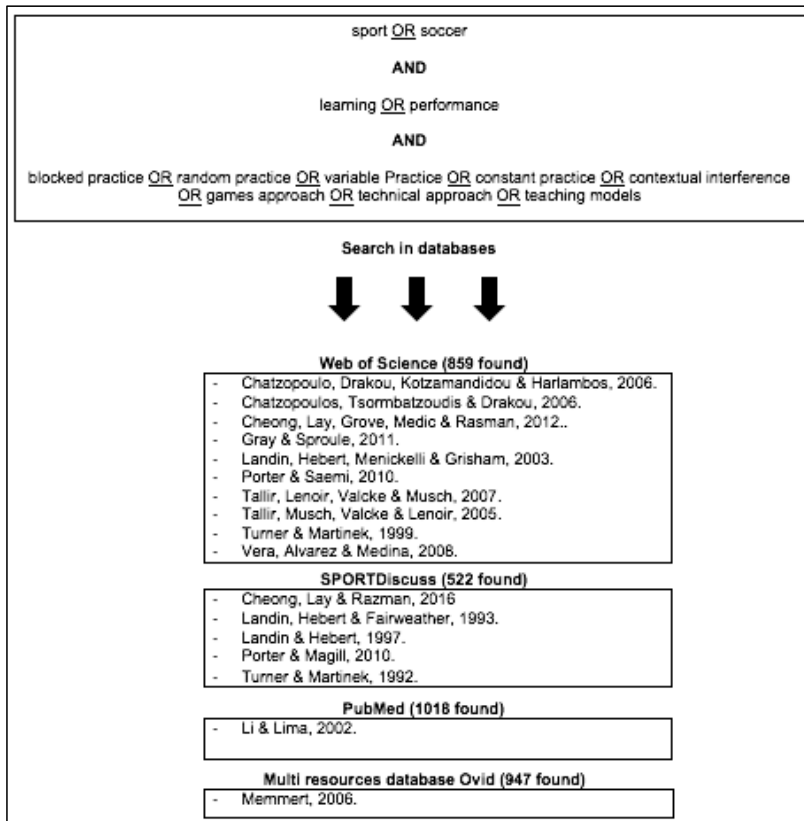


Figure 1: An in-detail overview of the search computed, including key word combinations, total number of studies found, and the final studies included in the review.

2.2 Inclusion/ exclusion criteria

The criteria for inclusion and exclusion studies used was: (a) studies had to be published in English in the period of 1990 and marc 2016; (b) they had to be original articles, not books, papers, thesis etc.; (c) full text had to be available; (d) all studies had to include a population from either physical activity, physical education or sport; (e) studies had to accumulate empirical data; (f) populations had to be divided in two or more treatment groups exerting the same sport, and could not contain subjects with challenges like eating disorders, psychological or physical illness; (g) studies had to

measure acquisition, through testing both pre and post of the treatment period; (h) subjects had to be exposed to invasion sports; (i) all journals publishing the studies had to be approved as a level one or two journal in the database of publication recommendations by NSD (NSD, 2015).

Clearly defined inclusion and exclusion criteria were important, with the objective of reliably determining which studies that fitted both the scientific aim of this thesis and had sufficient methodical quality. The criteria were chosen after reviewing similar work with similar methods in the domain of sport (Harwood, Keegan, Smith & Raine, 2015; Goodger, Gorely, Lavalee, Harwood, 2007; Van Rossum, 1990). For a more in depth overview of the whole process, when finding studies matching inclusion and exclusion criteria's, see Figure 2.

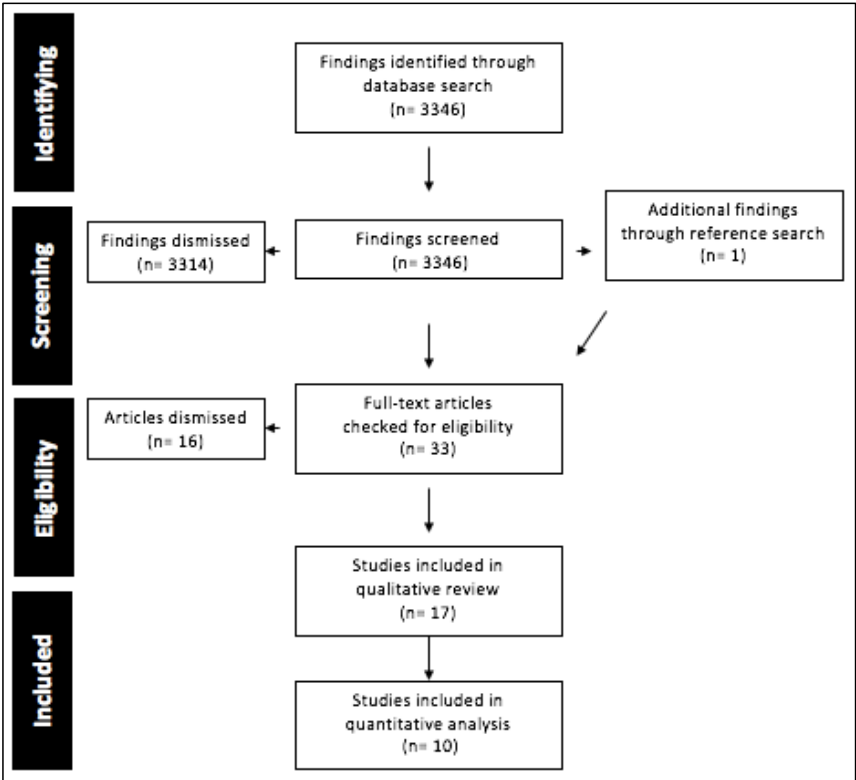


Figure 2: An overview of the inclusion and exclusion process.

Examples of studies excluded as they were seen as failing one or more criteria: (c) Cheong and Lay (2013); Cheong, Lay, Grove and Razman (2010); (f) Memmert and Roth (2007) tested across sports; Mesquita & Farias (2012) included only on intervention group; (g) Olosova and Zapletalova (2015) did not contain a pre-test before the treatment period; Berry, Abernethy and Cote (2008) was an retrospective study; (h) three studies were excluded due to testing outcomes which are so called especial skills, penalty throw, free-kick and throw inn (Breslin, Steenson & Williams, 2012; Savelsbergh, Canal-Bruland & Van Der Kamp, 2012; Tillar & Marques, 2013); (i) The journal Conte, Moreno-Murcia, Perez & Iglesias (2013) was published in were not approved; four studies was not found in the database (Allison & Thorpe, 1997; Priklerova & Kucharik, 2015; Psotta & Martin, 2011 Zuffova & Zapletalova, 2015).

2.3 Data extraction

Data included: (a) author; (b) year of publication; (c) country in which the main author works; (d) subject characteristics (N, age and gender, level); (e) Practice schedule (number, extent, stability); (f) evaluation characteristics (testing procedure, time point); (g) evaluation of learning effect. When details about average age of the study population were not given in a study, an average was estimated from the range of ages given (Harwood et al., 2015).

2.4 Analysis

2.4.1 Systematic review

As the selection process was over, 17 studies were included, and were reviewed. First, the relevant sections were read with the objective of getting familiar with it, and data was coded and gathered systematically. Secondly, similar data with different reported terms were clustered in to similar categories. For instance, the different terms named in the studies categorizing similar types of treatments (practice schedules) were clustered together. An overview is shown in Table 1.

Table 1: Included studies, categorized by type of practice schedule and degree of contextual interference

ID	Study	Higher	Moderate	Lower
1	Chatzopoulos et al., 2006a	Games approach	-	Technical approach
2	Chatzopoulos et al., 2006b	Games approach	-	Technical approach
3	Cheong et al., 2012.	Random practice	Mixed	Blocked practice
4	Cheong et al., 2016.	Game based	Random	Blocked practice
5	Gray & Sproule., 2011.	Games approach	-	Skill based approach
6	Landin & Hebert., 1997.	High CI	Moderat CI	Low CI
7	Landin et al., 2003.	High CI	Moderat CI	Low CI
8	Landin et al., 1993.	Variable practice	-	Constant practice
9	Li & Lima, 2002.	Random practice	-	Blocked practice
10	Memmert, 2006.	Variable practice	-	Constant practice
11	Porter & Magill, 2010.	Random schedule	Increasing schedule	Blocked schedule
12	Porter & Saemi, 2010.	Random schedue	Increasing schedule	Blocked schedule
13	Tallir et al., 2007.	IGC-model	-	Traditional approach
14	Tallir et al., 2005.	IGC-model	-	Traditional approach
15	Turner & Martinek, 1999.	TGfU	-	Traditional approach
16	Turner & Martinek, 1992.	Games approach	-	Traditional approach
17	Vera et al., 2008.	Variable	Combined	Blocked

Note: Higher= higher degree of CI. Moderate= moderate degree of CI. Lower=lower degree of CI.

Thirdly, the descriptive data were shown for each study individually in tables, and trends and patterns were described in text format. Last, results not fitting to quantitative methods were reviewed in a qualitative manner, by simultaneous comparison of the three groups.

2.4.2 Meta-analysis

In cases, like in this thesis, when the studies use different testing procedures to assess outcomes, the scale of the used measurements will vary, making it less meaningful to combine the raw mean differences. In such instances, it is recommended to use methods like Cohen's d (1988), as it creates a standardized difference in mean to be compared (Borenstein et al., 2009). Standardized differences in mean, the method of Cohen d , was used in this thesis to measure the magnitude of the effect sizes (Cohen, 1988). It estimates the differences, specifically the distance, between two groups and is calculated from the groups mean and standard deviation (Cooper, 2010). In all

calculations raw scores were used. As a recommendation from Cooper (2010), studies with three treatment groups were analysed by comparing two of the groups at a time.

Independent group post design (IGP) and independent group pre to post design (IGPP) were used when measuring outcomes in the studies included. IGP design compares different treatments of independent groups, who are tested at a single point in time, after the intervention (Morris & DeShon, 2002). Although the outcome measure did not include pre-test measures, all studies included a pre-test, as this was an inclusion criteria of the review in this thesis. When outcomes derive from an IGPP design, it is measured both before and after treatment (Morris & DeShon, 2002).

When results were found using an IGP design, the post-test means (M_{post}) were subtracted in the results direction, and then divided by a pooled standard deviation ($SD_{post,P}$) of the groups post-test (Cooper, 2010, p.164; Morris & DeShon, 2002, p. 107). The effect size of results deriving from IGP can be estimated as follows:

$$d_{IGP} = \frac{M_{post,1} - M_{post,2}}{SD_{post,P}}$$

The pooled standard deviation was estimated by considering both the standard deviation (SD) and sample size (n) of both groups (Cooper, 2010, p. 164). To estimate the common standard deviation, the following formula was used:

$$SD_{post,P} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

When studies did not report mean and standard deviation, t-test results (t_{IGP}) were used to estimate the effect size (Cooper, 2010), using the method of Glass and colleagues (1981, found in Morris & DeShon, 2002), as follows:

$$d_{IGP} = t_{IGP} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

In cases when results derived from an IGPP design, an effect size from each group was first estimated to find the increase pre to post-test, subtracting pre-test mean (M_{pre}) from post-test mean ($M_{post,1}$), and then dividing it on the groups pre-test standard deviation (SD_{pre}). When the increase of each group was found, the two groups' effect sizes was subtracted in the direction of the significant results (Becker, 1988). The following formula was used (Morris & DeShon, 2002, p.108):

$$d_{IGPP} = \frac{(M_{post,1} - M_{pre,1})}{SD_{pre,1}} - \frac{(M_{post,2} - M_{pre,2})}{SD_{pre,2}}$$

When effect sizes from studies contained sample sizes with 20 subjects or less, it was corrected using Hedges (1980) to avoid sample size bias. For effects deriving from studies in the IGP design, studies with 20 subjects or less across the two groups of comparison was corrected for. As for effects from studies with an IGPP design, 20 subjects or less in one group was corrected, as an effect of increase inside the group gave the first estimate of an effect. The effect of d unbiased (d^U) is found using a correction factor (C_m), when the standard deviation has m degrees of freedom. The m degrees of freedom is an estimate of one less than the sample size (Hedges, 1980).

$$d^U = C_m d$$

As noted by Cooper (2010) there are no permanent rules for choosing a model of effect in meta-analysis, though are there recommendations. A random-effects model was used in this thesis, as there was not enough in common to make the assumption that the true effect size is identical across studies (Borenstein, Hedges, Higgins, Rothstein, 2010), as well as it seemed most fitting when the objective is to give implications in a broader sense in the area of soccer development. The two most used statistical models are the one chosen in this thesis and the fixed-effect model.

The models differ inter alia by how they define the total error variance (Borenstein et al., 2010), giving enough backdrop to choose the Random-effects model in this thesis. Since the variance of the effect, when comparing the studies, will differ between studies as well as between subjects in a study, there are more than one source of variance (Borenstein et al., 2010), supporting the choice of Random-effects model. This

backdrop is further described in the section of strengths and weaknesses, also giving another example. When using a random-effects model, the effect size estimated will be evaluated, and representative, as a random selection of effect in a domain. The goal is to estimate the average effect in the current domain. Fixed effects-model, make the assumption that the found effect is the one true effect (Borenstein et al., 2010), since the studies' virtue are samples of the total population (Shadish & Haddock, 1994).

Effect sizes from the two designs, IGP and IGPP, were separated based on two notions: (1) that the variance of score in the IGP design was not homogenous across time, and will by this separation give less biased and more precise measures (Morris & DeShon, 2002); (2) systematic differences in effect sizes was found between the two designs, when testing mean differences (Morris & DeShon, 2002).

When combining effect sizes in one of the designs two methods were used. One when computing an overall effect of all studies, as an estimate where each study was an individual unit of effect size, as this is the recommended method. Although, in the face of a general low quantity of studies available in this thesis, another option of method was chosen for the remainder of the results. When computing by the other method, each outcome was treated as an individual unit, giving several effect sizes from a single study in some instances. In the results it is stated when each of the methods is used. These choices are discussed in the section of strengths and weaknesses.

Comprehensive meta-analysis (CMA), a computer program for meta-analysis, was used when merging the single effect sizes. The estimated standardized difference in means and variance was inserted, after computing them as described above. As well as each treatment group's sample size and the effect size direction. Some studies, where the number of subjects in each treatment group were not specified, had to be calculated from the total number of subjects. The total number of treatment groups was divided by the number of treatment groups.

Since the random-effects model was chosen, the standard error (*SE*) of summary effect is given by also taking into account the between studies variance (Borenstein et al., 2009). When estimating the variance of *d*, the uncertainty both of the mean differences and the standard deviation was accounted for, as the numerator and denominator in the

equation (Borenstein et al., 2009; Shadish & Haddock, 1994). This is a very close approximation of the d -index variance (Cooper, 2010). The standard error of d is the square root of the variance (Borenstein et al., 2009).

The criteria for classification of the effects magnitude was as following, recommended for behavioural sciences and for testing between independent means (Cohen, 1988; Cohen, 1992); .20-.49 small (S), .50-.79 moderate (M), and .80 and above represents large (L) differences. A Confidence interval (CI), of each estimate (k), was constructed around the scores of d (Cooper, 2010, p.14) with a chosen level of CI of 95% ($CI_{d,95\%}$). The estimation was calculated as follows, where CI is the square root of 1, divided on the sum of the studies' weighting factors (w_i):

$$CI_{d,95\%} = d. \pm \sqrt{\frac{1}{\sum_{i=1}^k w_i}}$$

The weighting factor, is extracted in CMA, but was additionally estimated manually for use when estimating the mentioned CI, and as later described Q statistics. Weighting factor, in relation to each effect size, is the inverse of the variance (v_i) (Cooper, 2010), Estimated as 1, divided on the variance (Shadish & Haddock, 1994, p. 265):

$$w_i = \frac{1}{v_i}$$

The variance can be computed, by using estimates of the studies effect size (d) and sample sizes of the two groups (n), as recommended by Cooper (2010, p. 165) and Shadish & Haddock (1994, p. 268):

$$v_i = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}$$

The homogeneity of the effects was tested through a significance test, named Q -statistics (Hedges & Olkin, 1985). Homogeneity analysis, not only compare the expected variance from sampling error to the actual observed variance. It also calculates how likely it is that, the variance making the studies different solely is a result of

sampling error (Cooper, 2010). Testing if the set of d effects is homogeneous by calculating the statistics of Q_t (Cooper, 2010, p.186):

$$Q_t = \sum_{i=1}^k w_i d_i^2 - \frac{(\sum_{i=1}^k w_i d_i)^2}{\sum_{i=1}^k w_i}$$

The formula above gives a value, which is compared to a table of chi-square values, specifically upper tail (Cooper, 2010, p.188). The chi-square value is $k - 1$ degrees of freedom, determined by number (k) of d effects, minus one (Cooper, 2010; Morris & DeShon, 2002). The chosen level of significance was .05 ($p < .05$) in this thesis across all tests of homogeneity. If the obtained value of Q_t was lower than the critical value found in the table at the chosen significance level, the hypothesis that the variance of effects was by sampling error alone is confirmed (Cooper, 2010).

3. Results

3.1 General results

An overview, showing all studies, is computed in Table 2, where all experiments are based on empirical evidence ($p < 0.05$) of learning effects. In total, 17 studies had a total population of 1019.

The variation of age reported in the studies population was between 9.5 and 24.6 years, in which 47.1% of the studies had an reported average age ranging from 18 to 24.6 years old (adult), and 51% from 8 to 12.5 years old (youth). No studies reported an average age in between. Two studies (11,8%) did not report age numerically. Most studies include both males and females in the study populations ($k = 9$, 52.9%). The remaining studies had only male ($k = 3$, 17.7%), female ($k = 2$, 11.8%) or did not report details of gender in the population ($k = 3$, 17.7%). In general, in the studies reporting level of expertise prior to the treatment period ($k = 11$, 64.7%), they stated a low level. The remaining six studies (35.3%) did not report these levels explicitly. Of the eleven studies reporting, six of the studies subjects had no or little prior experience (54.5%), and two reported subjects to be beginners, implying very low level of expertise. The remaining three reported a step above, respectively medium ($k = 1$), moderate ($k = 1$), or 2 years' prior experience ($k = 1$). No studies reported subjects with high level of expertise.

Among the studies, seven countries of the authors' workplaces were reported. Most authors came from the USA ($k = 8$, 47.1%), the rest from Australia ($k = 2$), Belgium ($k = 2$), Greece ($k = 2$), Germany ($k = 1$), Scotland ($k = 1$) and Spain ($k = 1$, 5.9%). The proliferation of sports was mainly spread across the three sports Bandy ($k = 4$, 23.5%), basketball ($k = 8$, 47.1%) and soccer ($k = 4$, 23.5%), but also from Ice hockey ($k = 1$) and the sport of Ultimate Frisbee ($k = 1$).

Table 2: A descriptive overview of study and subject characteristics.

ID	Author	Publ.	Nationality	Sport	N	A	G	Level
1	Chatz. (a)	2006	Greece	Soccer	72	12.5	f	(middle school)
2	Chatz. (b)	2006	Greece	Soccer	101	12.5	-	(middle school)
3	Cheong et al.	2012	Australia	Bandy	55	18	m/f	No prior exp.
4	Cheong et al.	2016	Australia	Bandy	70	21.6	m/f	No prior exp.
5	Gray & Sproule	2011	Scotland	Basketball	52	12.5	m/f	(Physic educ.)
6	Landin et al.	1997	USA	Basketball	30	20.5	m/f	Two years exp.
7	Landin et al.	1993	USA	Ultimate	28	20.6	f	No prior exp.
8	Landin et al.	2003	USA	Basketball	34	20.3	m	No prior exp.
9	Li & Lima	2002	USA	Soccer	38	22.4	-	Beginners
10	Memmert	2006	Germany	Basketball	32	24.6	m/f	(Physic educ.)
11	Porter & Magill	2010	USA	Basketball	96	-*	f	Beginners
12	Porter & Saemi	2010	USA	Basketball	45	-*	m	moderate
13	Tallir et al.	2007	Belgium	Basketball	97	10.5	-	Little or non exp.
14	Tallir et al.	2005	Belgium	Basketball	97	10.5	m/f	(primary school)
15	Turner et al.	1992	USA	Bandy	44	12	m/f	No prior exp.
16	Turner et al.	1992	USA	Bandy	71	12	m/f	(6 th /7 th graders)
17	Vera et al.	2008	Spain	Soccer	57	9.5	m/f	Medium

Note: N= sample size. A= average age. G=gender. Level= level of expertise.

Details about the treatment, testing and significant differences are shown in Table 3. Among the studies, only one did not report the number of treatments in the intervention. Several ($k = 7$, 41.2%) reported between one and five treatments, but most reported between eight and 15 treatments ($k = 9$, 52.9%). The treatments contained practice stated as trials in seven studies (41.2%), and as minutes in eight studies (47.1%). Two did not report specific amounts of practice.

Overall, among the 17 studies, 12 of them included pre to post intervention test (70.6%), seven testing at a retention time point, five testing of transfer effect (29.4%) and three did assessment test during the treatment period (17.6%). Five of the studies did not report neither significant effect, nor positive or negative to the degree of CI. Among the 12 studies that did ($p < .05$ or lower), five showed significant results on more than one type of test (41.7%). Studies with significant differences in learning effect mostly used pre to post intervention tests, especially eleven studies. That is, eleven out of the twelve which conducted this testing procedure (91.7%). But also the other test

time points showed at least one test with significant differences out of the total conducted, five out of seven at retention (71.4%), two out of five at transfer (40%), and one out of three during treatment (33.3%), respectively.

Table 3: Describes the key features relevant in the treatment and testing procedure, who are common across groups, as well as significant differences in learning between groups.

ID	Author	Publ.	k	T/M	Evaluation	Learning
1	Chatz., Drako.(a)	2006	15	45 m	a	a=+H
2	Chatz., Tsorm.(b)	2006	15	45 m	a	a=+H
3	Cheong et al.	2012	8	45 t	a, b, d	-
4	Cheong et al.	2016	10	15 m	a, d, e	a=+H.e= +H*
5	Gray & Sproule	2011	(5)	(PEC)	a	a=+H
6	Landin et al.	1997	3	30 t	a	a=/*
7	Landin et al.	1993	3	120 t	b, d, e	-
8	Landin et al.	2003	2	100 t	a	-
9	Li & Lima	2002	1	45 m	a, d	-
10	Memmert	2006	1	160 t	a, d, e	a=+L. d= +H
11	Porter & Magill	2010	1	81 t	d, e	d, e=/*. d,
12	Porter & Saemi	2010	5	81 t	a, d	d=/*, a=/ a=+H
13	Tallir et al.	2007	12	(PEC)	a, b, d	a, d=+H
14	Tallir et al.	2005	12	50 m	b	b=+H
15	Turner et al.	1992	10	35 m	a	-
16	Turner et al.	1999	15	45 m	d, c	d, c=+H, c*
17	Vera et al.	2008	15	50 m	a, d, e	a=/*

Note:

k= number of treatments (when not stated, weeks is shown in parenthesis).

T/M= extent of each treatment given in T; trials, or M; minutes. (when not stated, physical education class is shown in parenthesis)

Evaluation= method of evaluation (a; pretest-post-test. b; assessment test during treatment. c; control group test. d; retention test. e; transfer test)

Learning= significant learning differences ($p < .05$ or lower)

+H= high degree of CI > low degree of CI.

+H*= high degree of CI > low and moderate degree of CI.

+L= low degree of CI > high degree of CI.

/= moderate degree of CI > lower degree of CI.

/*= moderate degree of CI > higher or/ and lower degree of CI

c*= higher degree of CI > control group

-= No significant learning effect between groups

3.1.1 Studies with three treatment groups

The results, when looking only at studies containing three groups, are reviewed in a qualitative manner as well. The chosen method of the meta-analysis did not contain comparisons of three results simultaneously, as recommended by (2010), and are therefore being compared in pairs in the quantitative section. In that section the moderate degree of CI groups is categorized as high when compared to low degree, and as low when compared to high degree. Here, in this section, each group is treated according to their descriptive nature but in a qualitative way to shed light on moderate degrees of CI.

The number of studies with three groups with significant results are six (Cheong et al., 2016; Landin & Hebert, 1997; Porter & Magill, 2010; Porter & Saemi, 2010; Turner & Martinek, 1999; Vera, Alvarez & Medina, 2008), though one of them did not include a practice schedule with a moderate degree of CI, rather a control group (Turner & Martinek, 1999), making the study illegible for the results in this section. As shown in Table 4, most of them are in favour of moderate degree of CI compared to high or/and low degree. More specifically, there are 18 significant results ($p < .05$ or lower) in the five studies.

When high degree of CI was compared to moderate, six results from four studies were in favour of moderate degree of CI. All results were from studies with skill tests as testing procedure. Only two results, in a single study, contradict the direction favouring a higher degree (Cheong et al., 2016). One of them deriving from a skill test. The last result was from a gameplay testing procedure, in fact the only game play test done in the studies with three groups. The groups treated in the studies differ with regard to stability of environment under the intervention. Further details about results in the studies are shown in Table 4.

Table 4: Stability of environment in treatments of high and moderate degree of contextual interference in studies with three groups. Quantity, outcome and direction of the significant differences between groups.

ID	Author	Moderate	High	k	Outcome(s)	M vs H
1	Cheong et al.	Stable	Unstable	2	Execution (passing)	<
2	Landin &	Stable	Stable	1	Execution (shooting)	>
3	Porter &	Stable	Stable	2	Execution (passing)	>
4	Porter &	Stable	Stable	1	Execution (passing)	>
5	Vera et al.	Unstable	Unstable	2	Execution (shooting,	>

Note: Moderate= Stability of environment (moderate CI). High= Stability of environment (high CI). K= number of significant results. Outcome(s)= skill. M vs H= Direction of significant results.

Both Moderate and high degree of CI was favoured in all comparisons to lower degree of CI, respectively seven results in four studies in favour of moderate, and four results in three studies favouring higher degree of CI. 5.2 Meta analysis

3.2 Main results - Meta analysis

12 studies met the criteria for inclusion and showed significant ($p < .05$ or lower) results. Two studies (Porter & Magill, 2010; Turner & Martinek, 1999) and one effect size in a study (Porter & Saemi, 2010) were omitted because of insufficient data. However, the 10 remaining studies offered 25 effect sizes when combining the independent groups pre-test to post-test design ($k = 16$) and independent groups' post-test ($k = 9$) estimates.

Treating the studies as an individual unit of analysis, where effect sizes are calculated as an average of all estimates in a study, gave an average effect size of .91 ($SE = .20$, $CI_{.95\%} = .22$) deriving from the estimates with IGPP design, and .38 ($SE = .22$, $CI_{.95\%} = .30$) for the IGP effect sizes. Both IGPP [$Q(6) = 10.77, p < .05$] and IGP, [$Q(3) = 5.62, p < .05$] designs averaged effect sizes which significant tests for homogeneity, confirming the hypothesis that sampling error was the lone reason of variance across the four effect sizes. One study was included in both averaged effect sizes (Cheong et al., 2016). When excluding studies that reported three different practice schedules, an effect size averaged was reported of .97 ($SE = .20$, $CI_{.95\%} = 0.25$) for the IGPP designs, and another significant homogeneity test [$Q(4) = 8.5, p < .05$].

For the remaining results presented, each outcome effect size was treated as an individual unit, also when they were an estimate from a study with three practice groups with different schedules. IGPP had an overall mean effect size of .73 ($SE=.19$, $CI_{.95\%}=.15$), and a insignificant homogeneity test [$Q(15)=101.04, p<.05$]. The H-statistic value is greater than the critical value of significance, and rejects the hypothesis that sampling error was the only factor of variance in effect sizes. IGP designs reported a slightly lower overall effect size, .51 ($SE=.27, CI_{.95\%}=.20$), and once more an insignificant homogeneity test [$Q(8)=47.55, p<.05$], respectively. The estimates of effect sizes and standard error for potential moderator groups are included in Table 5.

Table 5: An overview of moderating variables effects, and their estimates (quantity of individual effect sizes, standard error, confidence interval, classification of magnitude and homogeneity).

Variable	IGPP	IGPP	IGPP	IGPP	IGPP	IGPP	IGP	IGP	IGP	IGP	IGP	IGP
Age	N	ES	SE	CI	Cl.	Hom.	N	ES	SE	CI	Cl.	Hom.
Youth	12	.71	.23	.18	M	N	2	-	-	-	-	-
Adult	4	.79	.41	.34	M	N	4	.39	.35	.35	S	N
Outcome	N	ES	SE	CI	Cl.	Hom.	N	ES	SE	CI	Cl.	Hom.
Decision making	4	.80	.16	.26	L	Y	1	-	-	-	-	-
Execution	9	.39	.26	.21	S	N	7	.42	.40	.28	S	N
Recognition	-	-	-	-	-	-	1	-	-	-	-	-
Tactical	1	-	-	-	-	-	1	-	-	-	-	-
Support	2	-	-	-	-	-	2	-	-	-	-	-
Stability of	N	ES	SE	CI	Cl.	Hom.	N	ES	SE	CI	Cl.	Hom.
Stable	2	-	-	-	-	-	5	.32	.64	.38	S	N
Unstable	9	.81	.30	.18	L	N	2	-	-	-	-	-
Unstable vs Stable	5	.81	.27	.28	L	N	2	-	-	-	-	-
Testing	N	ES	SE	CI	Cl.	Hom.	N	ES	SE	CI	Cl.	Hom.
Game play	6	1.26	.33	.22	L	N	2	-	-	-	-	-
Skill test	9	.39	.26	.21	S	N	6	.40	.50	.32	S	N
Video	-	-	-	-	-	-	1	-	-	-	-	-
Questionnaire	1	-	-	-	-	-	-	-	-	-	-	-
Testing time	N	ES	SE	CI	Cl.	Hom.	N	ES	SE	CI	Cl.	Hom.
Assessment	-	-	-	-	-	-	1	-	-	-	-	-
Post	15	.73	.21	.15	M	N	5	.55	.28	.28	M	N
Retention	1	-	-	-	-	-	3	.41	.92	.42	S	N

Note: N= number of effect sizes. ES= Total effect size. SE= standard error. CI= confidence interval. Cl.= classification of effects magnitude (L= large, M= moderate, S= small). Hom.= Homogeneity (Y= yes. N=no)

IGPP, which takes in to account the pre-test to post-test difference, was used when describing further results in more detail. When the analysis by this design did not provide a sufficient number of estimates, effects sizes summed from the IGP design were used, when available.

Since execution is used as a common term for all execution skills in the studies, as opposed to the same descriptive nature across studies (e.g. decision making), execution

as a variable need further elaboration. Although it has to be noted that also decision making differed slightly. One of the studies included both execution and decision making (Gray & Sproule, 2011). The ability to make appropriate choices about what to do in a situation with the ball is the definition of decision making in the remaining three studies with significant decision making differences (Chatzopoulos, Drakou, Kotzamandidou, Harlambos, 2006).

The effect sizes and standard error of the various types of execution skills as a moderator are shown in Table 6.

Table 6: Overview of the execution outcomes specifics, with their effects, and their estimates (quantity of individual effect sizes, standard error, confidence interval, classification of magnitude and homogeneity).

Variable						
Execution skills	<i>N</i>	<i>ES</i>	<i>SE</i>	<i>CI</i>	<i>Cl.</i>	<i>Hom.</i>
Dribbling	2	-	-	-	-	-
Passing	5	.55*	.49	.31	M	N
Shooting	5	.12	.19	-	X	-

Note: *N*= number of effect sizes. *ES*= Total effect size. *SE*= standard error. *CI*= confidence interval. *Cl.*= classification of effects magnitude (L= large, M= moderate, S= small). *Hom.*= Homogeneity (Y= yes. N=no). *=Learning effect measured with IGP design

Combining the variables testing procedure and stability of environment are estimates of effect sizes which are more or less like the actual game, depending on the variables. Effect sizes and standard error as a combined moderator are show in Table 7.

Table 7: Effects of different stabilities in case of the environment in the treatment and testing procedures,, and their estimates (quantity of individual effect sizes, standard error, confidence interval, classification of magnitude and homogeneity).

Variables							
Stability of	Testing procedure	<i>N</i>	<i>ES</i>	<i>SE</i>	<i>CI</i>	<i>Cl.</i>	<i>Hom.</i>
Stable	Game play	0	-	-	-	-	-
Stable	Skill test	5	.32*	.64	.38	S	N
Unstable	Game play	6	1.26	.33	.22	L	N
Unstable	Skill test	2	-	-	-	-	-
Unstable vs stable	Game play	1	-	-	-	-	-
Unstable vs stable	Skill test	5	.81	.27	.28	L	N

Note: *N*= number of effect sizes. *ES*= Total effect size. *SE*= standard error. *CI*= confidence interval. *Cl.*= classification of effects magnitude (L= large, M= moderate, S= small). *Hom.*= Homogeneity (Y= yes. N=no) *=Learning effect measured with IGP design

In the case of combining stable environments and testing using a skill test, all estimates were found with execution as an outcome. Of the five estimates, three were post treatment, and the remaining two were done at a retention time point. The execution outcome differed in case of type of skill, respectively passing ($n = 3$) and shooting ($n = 2$). The estimates from the execution of passing outcomes had an effect size of .48 ($SE = 1$, $CI_{.95\%} = .47$). H-statistics showed an insignificant test for homogeneity across studies [$Q(3) = 34.89$, $p < .05$] rejecting the hypothesis that the variance in the effect sizes was due to sampling error alone.

Similarly, as in the case of solely stable environments, execution was the significant outcome when combining a group in an unstable environment with a stable one. In this case, all the skill tests were done at a post time point. The various skills performed in the execution outcome were dribbling ($n = 1$), passing ($n = 2$) and shooting ($n = 2$).

Estimates when both groups were exposed to an unstable environment and later tested in game play, were reports of both support ($n = 2$) and decision making ($n = 4$) as outcomes. All tests were at a post time point. Decision making alone had an average effect size of .80 ($SE = .16$, $CI_{.95\%} = .26$), and a significant homogeneity test [$Q(3) = 4.02$, $p < .05$].

3.2.1 Correlates below threshold

Some moderating or sub- groups did not pass the threshold of three individual units with significant results. This applied to the outcome variable of support, which only contained two effect sizes. Interestingly, both were large single effect sizes, and one as high as 3.51.

The highest effect size was a significant contributor to the variance of all the effect sizes, when each outcome was treated as an individual unit within the IGPP design. The other two outcome variables below the threshold with only one unit each, was recognition, the single unit tested and found significant when subjects tested were watching video clips, and tactical knowledge, the only one using a questionnaire. When splitting the outcome variable of execution in to more specified estimates, dribbling fell below the threshold containing only two individual units.

In case of testing procedure, assessments during intervention period did only find one unit with significant results, and skill test procedure together with an unstable environment of treatment contained only two individual units.

4. Discussion

This master thesis aimed to enlighten how contextual interference affects learning of expertise in soccer. A systematic review, which also included quantitative measures of magnitude through a meta-analytic approach, was chosen to show the existing available research. In total 17 studies, dated back to 1990 and concluded in March 2016, were found appropriate and explored. The learning effects include both support of earlier research, contradictive support and new directions. These effects are briefly shown in the summary, and are discussed in detail later, under the section of learning effects. It has to be noted that most of the outcomes have methodical implications challenging their true meaning of the results. These implications are briefly named when presenting the effects, and are discussed in greater detail in the section of strengths and weaknesses. Ending this thesis, a conclusion is described, mainly focusing on implications for future practice and research.

4.1 Summary of learning effects

Overall, 12 of the 17 studies had significant effect one way or the other. Most were exclusively in favour of a high degree of CI, but some also included significant results in favour of a low degree. The effect supports the general understanding of the Contextual Interference effect theoretical framework (Battig, 1966; Brady, 2006), where high degree of CI and variability favour a learning effect.

Only one study reported a group, stated as low degree of CI, with significantly better effect than a group with higher degree of CI. The test was taken post treatment, and the effect was reversed in the same study at a retention time point, then supporting high degree of CI. This reversed effect is in support of the original notion (Battig, 1966) that random and variable practice will be detrimental for performance short term, but superior long term in case of learning. Though, in total, when including all studies both tests at a post and retention time point effects were in favour of high degree of CI, with moderate and small effect, respectively. In other words, a higher effect was achieved when testing post than at a retention time point.

Many of the moderating variables tested in the meta-analysis did not report differences. Youth and adult subjects both showed moderate effect, in fact, almost a large effect sizes at .79 and .71. As long as at least one group was exposed to a practice schedule with an unstable environment, it favoured higher degree of CI with a large effect. Not enough effect sizes were reported in stable environments deriving from the IGPP design as above. The IGP found a small effect in favour of high degree of CI, for practice schedules containing solely stable environments.

Although the effects were mostly in favour of high degree of CI, the magnitude of some of the moderating variables varied. A stronger relationship was found for decision making than execution, large and small effect, respectively. Across execution outcomes only passing and shooting had strong enough effect sizes to be compared, and only passing was found having an effect above .20, which is the threshold for small effect. Specifically passing had a moderate effect. When testing in game play the largest effect was found of all estimates in the analysis, with an effect size of 1.23. Skill tests in comparison did only report a small effect.

A moderate degree of CI was presented in the qualitative analysis, where various results were shown compared to high degree, mostly in favour of moderate degree. All testing procedures in favour of moderate degree derived from skill tests, and only one skill test contradicted this trend by being in favour of high degree of CI. The only game play testing procedure, also comparing moderate and high degree of CI, was in favour of high degree. Both moderate and high degree of CI were superior to low degree in all significance results from studies containing three groups.

4.2 Learning effects

4.2.1 Overall

Twelve studies found significant results in one of the directions, and mostly in the direction of high degree of CI. Five studies could not report significant results, but most of these reported trends, also in favour of high degree of CI. As this master thesis tested the true meaning of the results, the magnitude of the relationship. When studies were treated as individual units in an independent group pre to post-test design, a large effect was found in favour of high degree of CI. The same large effect was found when

excluding studies with three practice groups. These effects indicate, as reported earlier (Battig, 1966; Brady, 2008), the benefits of practicing in a randomized and variable manner if learning is the goal.

In fact, the large effect reported overall exceeds similar meta-analysis with a broader motor research domain, including more than invasion sports, which have only found a small effect (Brady, 2004). Results from studies using an independent group post-test design found a small effect, in a similar manner to earlier and broader research.

Despite variations of effect in case of design, both indicate the benefits of high degrees of CI in sports researched in this thesis, primarily soccer, but also other invasion sports. Indeed, the large effect found from estimates in the IGPP designs tests has a lower standard error, confidence interval, both with and without studies including three groups, and is a better research design in general than the IGP. Both effects passed the homogeneity test. As a result of the methodical advantage, if one of the effects was chosen, the effect of IGPP is superior.

4.2.2 Stability of environment

In relation to the nature of soccer, as well as other invasion sports, the stability of environment is dissimilar to many other motor tasks. Soccer is a complex sport, including cognitive and perceptual skills, as well as motoric, in an ever-changing environment (Bate, 1996). The notion of greatest improvements when the movement patterns in practice replicate the demands of the game context (Rushall & Pyke, 1990), suggests an unstable environment when practicing soccer. This can also be said for other invasion games as described in detail earlier, how the distinctions between categories in sport, where soccer is an invasion game, is described by Griffin and Butler (2005), and the transfer between the sports is shown at least partly (Causer & Ford, 2014; Smeeton et al., 2004).

When exposing both treatment groups to unstable environments, distinguishing them by degree of CI, a large effect was found in favour of high degree. This effect implies more play in game context, rather than less. In the case of learning methods used, it favours TGFU (Thorpe et al., 1986) and the similar learning methods described in the section of

theoretical knowledge, respectively GCA (Griffin et al., 1997) and PP (Lauder, 2001), not TA approach (Wright et al., 2005).

A similar large effect was found in favour of high degree of CI, when one of the groups was exposed to a stable environment. This implies superiority of practicing in a context replicating the game, as noted by Rushall & Pyke (1990). The effects when both groups were treated in a stable environment did not reach the threshold of three individual units of effect size using IGPP like the other types of environment, so the five estimates found from IGP designs were used. These estimates concluded with a small effect in favour of high degree. Implying in practice, benefits of a more random and variable practice schedule also when allowing the learner with a self-paced tempo. Although a trend in favour of a moderate degree of CI when practicing in a stable environment has been noted, this trend was not found using quantitative estimates of magnitude. Rather, it was found in the qualitative reviewing process. This trend, and other results from the qualitative reviewing of these studies are described in further detail in the section of three treatment groups.

Just like the treatment practice schedule should be replicating the actual game context, naturally, the testing procedure should be representative of the game context. More about this, the validity and reliability of the testing procedures, will follow in the section of strengths and weaknesses.

As described in this section, when taking the nature of soccer in account, both the practice schedule and the testing procedure should replicate the soccer game environment which is rather unstable. When only including estimates of effects where both groups were treated and tested in an unstable environment, a large effect of 1.26 was found. This is larger than all the other computed effects, including the overall effect for the same design of .91. In relation to the increase in effect when moderating for only the estimates with highest degree of similarity, the CI effect on learning may be even more robust in soccer and invasion sports than in more general motor tasks domains.

4.2.3 Outcomes

Significant differences were found in a total of five outcomes in the included studies. Two categories of outcomes had enough estimates of single effect sizes, above the

threshold of quantity, to estimate a total effect of magnitude, decision making and execution, respectively. Support, recognition, and tactical knowledge follow below.

Decision making had a large effect, just above the threshold of large magnitude, with .80. This effect was found with a low standard error, compared to other effects in this study, and a significant homogeneity test, which facilitates the meaning of the effect. All effects were from different studies, in other words studies were the unit of analysis. For a more detailed elaboration on units of analysis, see the section of strengths and weaknesses.

The testing outcome of decision making in one of the studies, included both the decision making and the execution of the skill (Gray & Sproule, 2011). In support of this inclusion it has been noted that, the technical execution in competition elapses from the cognitive and perceptual, which forms the decisions (Williams, 2000; Williams & Hodges, 2005). The rest of the studies did only find significant results about the decisions of the individuals.

Decision making, as described earlier, is a distinguishing factor between levels in soccer (Kannekens et al., 2009; McMorris & Graydon, 1996), and has been shown transferable across invasion sports (Causer & Ford, 2014; Smeeton et al., 2004). With the importance of decision making, and large effect of magnitude in favour of high degree of CI in soccer, and across invasion sports, the implication of practicing in a random and varied manner is a robust finding. In retrospective studies this is supported, where the amount of time spent in activities containing decision making has been shown as a distinguishing factor between levels of expertise (Ward, Hodges, Starkes, Williams, 2007).

Another outcome with similar treatment and testing procedure as decision making was the skill of support. In total, these two skills have all the effect sizes reported from significant results in unstable environment and testing procedures. Support is reported in the studies as actions off the ball (Chatzopoulos et al., 2006a; Gray & Sproule, 2011). Solely two effect sizes were found, making the total effect below the inclusion threshold. Both single effects were favouring high degree of CI, with effects of .91 and

3.52, respectively. The last one is the highest single effect size found, regardless of outcome.

In cases of only moderate outcome, selecting execution, a small effect of magnitude was found in favour of a high degree of CI. When additional testing and treating of both groups in an unstable environment was performed, only one study demonstrated significant differences favouring high degree of CI (Gray & sproule, 2011), and that was the combined measure of the total decision making and execution in gameplay. This is also the only effect size found, when treatment and testing procedure both are in an unstable environment. The technical quality should not be tested outside the game context, where important constrains of adversity, like variability and randomness, are missing. These are important components affecting space and time when executing an action (Ali, 2011; French, Werner, Rink, Taylor & Hussey, 1996).

Only one study has reported two effects, significant execution differences in favour of moderate degree of CI, when compared to high (Vera et al., 2008). These were derived from skill tests. Further elaboration is found under the section of studies with three groups of comparison.

No other significant results in either direction were found, when testing execution outcomes, after treating and testing the subjects in an unstable environment. Several studies have tested in this manner, but no one has been able to find significant differences. For example, if we look at the three studies that have found significant differences between treatments in decision making (Chatzopoulos et al., 2006a; Chatzopoulos et al., 2006b; Tallir et al., 2007) they all found a significant increase with regard to execution in both groups. But only differences between groups in execution outcomes found were trends, which in fact diverged in both directions.

As a consequence of the fact that both groups increased the level of execution in similar manner, no matter the degree of CI, as long as they were in a practice schedule which was unstable. It may give an indication that the environment is a more specific predictor than the degree of CI, when choosing practice schedule.

When comparing a treatment group exposed to an unstable environment with a group of a stable one, also tested in an unstable environment, only one study found significant differences on execution in favour of high degree of CI, i.e. in dribbling. This is the only study that did this type of comparison (Cheong et al., 2016). Testing execution, with the use of skill test instead of game play, are more common. In total five estimates, from three studies, have shown a large effect in favour of a high degree of CI. Specifically, the execution outcomes with significant effects were passing and shooting times two and dribbling. In sum the results, when including testing in game play and skill tests, all tests were in favour of high degree of CI. This implies a practice schedule with an unstable environment over a stable one, in the case of learning execution skills.

Execution in treatments where both groups, including the test, was done in a stable environment was found to have a small effect in favour of high degree. The discussion of the degree of CI in a stable environment, and how it affects execution is discussed in the section of studies with three groups.

To sum up, both decision making and execution favour high degree of CI, but in varying degrees. Also other moderating variables as testing and treatment stability, contributes to the variations.

4.2.4 Testing time point

Battig's (1966) original notion proposed benefits in the case of performance with low degree of CI when testing directly post treatment, and a reverse benefit favouring high degree of CI when testing at a retention time point. Only one study showed support of both these effects post and retention (Memmert, 2006). This study is separated from most, by the length of the treatment period, only containing one treatment. Two more studies have the same number of treatments, of only one, but none of them report the same effect. One did not contain a post-test (Porter & Magill, 2010), so naturally this effect was not present, whereas the other did not find significant differences at all (Li & Lima, 2002).

In the conducted meta-analysis, post-test had a moderate effect and retention a small effect, when combining effect sizes from the IGP design. The retention test effect was estimated with only three single effect sizes, and had a very high standard error of near

1. Despite these methodological weaknesses, the retention effect is similar to those reported for a broader domain of motor tasks. The post-test effect of large is higher (Brady, 2004). The effect from the IGP designs support the IGPP estimates, with a similar large effect. When concluding in this section, the results seem equally favouring high degree of CI, independent of testing time point stated in the studies.

4.2.5 Age

As a matter of instability in perceptual and motor skills, when growing and maturing, age is a potential moderating variable on learning effects (Davids et al., 2012). The similar results found across age groups in this study, where both adult and youth had a moderate effect favouring high degree of CI, is contrary to both theoretical knowledge and effects found in a broader domain of motor tasks.

It has been suggested, that benefits of a practice schedule with high degree of CI for adults, may not give the same learning effect for youth (Guadagnoli, Holcomb & Weber, 1999). Earlier studies have supported this suggestion (Lin, Wu, Udompholkul & Knowlton, 2010). Also support from the meta-analysis on a broader motor domain than this paper, where young subjects did not even have a small effect of magnitude, and adults had a moderate effect (Brady, 2004). Brady (2004) suggest an explanation, that the thoughts behind this difference is that the high complexity may be too overwhelming for children, as they have a more limited ability to process information.

Despite Brady's (2004) reported effects, there has also been contradictive effect how CI affects learning at different ages. In a quantitative review, the benefit of high CI on learning decreased with aging (Yan, Thomas, Thomas, 1998). This is in line with earlier suggestions by Shapiro and Schmidt (1982), that because children are lacking motor experience compared to adults, they benefit from a more varied practice schedule.

In sum, earlier research has been inconsistent, and in this study age was not a modifier of the effect. If a direction has to be chosen, it supports the latest hypothesis of Yan and colleagues (1998), based on the small effect found estimated from the IGP-design. Though, it is not recommended comparing effects from the two designs.

4.2.6 Three groups of comparison

An interesting and inconsistent result, was found when looking at studies with three treatment groups exclusively, including a moderate degree of CI. The moderate degree was superior to the high degree in some of the results. Although the reviewing of three groups did not include effect of magnitude, so the truest meaning available, including a third group with moderate degree of CI, had to be done qualitatively. All significant results were found with execution as variable in the test.

Both moderate and high degree of CI were superior when significant results were found, when compared to a group with low degree. High degree of CI was also the significantly better practice schedule, compared to moderate degree, in the only game play test. Similarly, in the only study where a group was treated with an unstable environment and opposed to a stable one, a skill test showed benefits of high CI in the unstable environment. Both effects support a high degree of CI derived from the study of Cheong and colleagues (2016).

Opposed the general notion, that the higher degree of CI is equal to better learning outcomes, the moderate degree was found to be superior to high degree across significant results, in studies when both groups were exposed to stable environments. The skills with significant differences were passing times three and shooting. Although all were tested in a stable environment, specifically skill test, which is more similar to the practice schedule environment than the game context; this is contradictive to the other effects. Since there is no effect deriving from quantitative measures on how to facilitate execution skills when practicing in a stable environment and testing in game play, such an outcome offers implications both for practice and for research described in the last column in this section.

Similar effects, in favour of moderate degree, significant results were found in one study, when both treatments exposed to was in an unstable environment, and subjects were tested in skill tests. The two execution outcomes were shooting and dribbling. This is not in support to the effect found in the same environment in game play, but neither can be seen as contradictive. Nor did any of the significant results which estimated the effect sizes in game play include execution outcomes. Execution outcomes were tested, but differences were not found, only trends.

Implications for practice seems to be that, a moderate or increasing degree of CI in the practice schedule is more beneficial for learning than a high degree, when training in a stable environment. It has to be mentioned, however, that this at least partly is because of the lack of testing in a more unstable environment. In the case of research, the need of comparing several degrees of CI, tested in a more unstable environment, will give more robust implications for practice in support of the above or contradictive.

4.2.7 Lack of learning effects

In this thesis moderating variables were included and excluded based on criteria described earlier. The effects based on this selection of variables are described above. In this section the excluded variables are discussed, and relevant research are illuminated.

The level of expertise in the studies included in this paper did not contain explicitly high levels. More specific, none of the studies have a population deriving from a level of above moderate or medium, and half of the studies which reported levels of expertise were categorized as reflecting no or little experience. The low level of expertise reported, along with the fact that six of the 17 studies did not state any level at all, were the reasons for not selecting this as a moderating variable to be examined.

In relation to levels of expertise, Wulf and Schmidt (1994) have proposed that, random practice may result in to excessive variability, which may in turn may have a detrimental effect on the development. Even though level of expertise was not selected as a moderator in this study, as it solely includes medium levels and below, contradict the suggestion above. More research is needed to draw safe conclusions about the CI effect on levels of expertise, but above all the need is biggest at the highest levels.

As described in the section of general results, the quantity of studies which included only one gender was very low. Respectively, two solely with females, and three only including males. With this as backdrop, gender was not computed as a potential moderating variable. Personality not computed for the same reason. This was not even described in the general results, as only one study included personality differences (Tallir, Musch, Valcke & Lenoir, 2005). The current study found a positive relationship between decision making ability and a high degree of field independence as cognitive style.

4.3 Strengths and weaknesses

In research it is important to evaluate the steps taken by the scientist, showing the thesis' strengths and weaknesses. Strengths and weaknesses in this thesis, and when fitting also in the studies reviewed, will be discussed in this section. When discussing the credibility, both validity and reliability are key concepts (Thagaard, 2013).

Reliability refers to the degree of reproducibility. If the same research is conducted several times, the results should be the same, if assuming no other change. In quantitative research it is a measure, if the results are reliable often measures with tests. When testing, the estimation of the instrument has to be used empirically (Hassmen & Hassmen, 2008).

Validity refers to the authenticity of the study. In which degree the study in fact is able to capture what is intended. Validity is often divided in two separate terms, where the internal validity is most relevant in experimental research (Hassmen & Hassmen, 2008), therefore not that relevant for the validity of the conducted review in this thesis. The external validity refers to the degree of generalization (Hassmen & Hassmen, 2008), which is of great relevance in relation to the aim of finding the CI effect is supported, and can be applied in practice.

4.3.1 Search and selection process

Of the studies found in the database search, 3346 studies (see Figure 2), 17 were included based on exclusion and inclusion criteria. When initiating the process of screening and testing studies for eligibility, as well as coding, a decision was made of not including a second researcher, although it would facilitate the reliability of the thesis (Cooper, 2010). Rosenthal (1978) examined recording errors in 21 studies, and found a frequency of errors between 0% and 4,2%, and 64% of the errors favoured the hypothesis of the researcher. In an empirical examination of the number of unreliable coding made in a meta-analysis, it has been found near perfect similarity between three different researchers when coding 30 documents. Only one judgment out of 27 did not reach an agreement across coders of 80% (Stock, Okun, Haring, Miller, Kinney & Ceurvorst, 1982). In regard to the quality of a coder, it is best to employ people with

prior experience. If such individuals are unavailable the coder should have time to train, to overcome the limitations (Cooper, 2010).

Ideally, all studies with the chosen set of criteria available in the world, was also found and included in this thesis. This is rarely the case, “*Even with the advent of (and perhaps due to partly over-reliance on) electronic searching, it is likely that some studies which meet our criteria will escape our search and not be included in the analysis*” (Borenstein et al., 2009, p.278). The missing studies could be as the included ones, giving more information of the effects by narrowing the CI and giving more power to the tests, or if different, give our sample more bias, weakening the meaning of the results (Borenstein et al., 2009).

In relation to missing studies, the studies can be kept out of sight, as a matter of publicity. This is a general concern, and if true will give a biased published literature (Borenstein et al., 2009). The reasons for not publishing them are often, a lack of statistically significant results, or a lack of large enough effects (Borenstein et al., 2009). Especially in studies with small sample sizes, the produced bias in the magnitude will be large (Borenstein et al., 2009). This may be the case in this thesis, containing a total population of 1019 subjects across studies, which is relatively small compared to other reviews in different research domains. In comparison the study of Harwood and colleagues (2015), which used the same methods of a review amplified by a meta-analysis, has a total population size of 34156 subjects. This is a weakness of this thesis, and give the implications more studies needed to facilitate the meaning of the effects.

The notion that, one may not be sure of the quantity of studies conducted but never published, has been named the file drawer problem (Rosenthal, 1979). There has not been any final solution to the problem, but a formula for establishing boundaries, estimating the degree of damage to research conclusions, has been articulated (Rosenthal, 1979). The use of this formula in this thesis would have strengthened the conclusion, but in the face of priorities relative to time, this problem was not examined.

Most criteria were chosen on the basis of regular researching standards like studies being full text available, original articles, accumulation of data in English language and so on. Some criteria in an attempt to capture the essence in the aim of the study, like

two or more treatment groups enabling comparison, solely including invasion sport and open skills with an aim to capture the nature of soccer, and only including physical active subjects with the goal of studying the relevant population of aim. Others were chosen with the aim of strengthening the quality of the thesis, like the need of a pre-test.

The decision to include other sports than soccer was perceived as necessary as only four such studies were found. The transfer, as noted earlier, has been at least partly supportive of learning across invasion sports, although the support is found in decision making outcomes (Causser & Ford, 2014; Smeeton et al., 2004), not for execution outcomes. Execution in soccer is, somewhat different than in other invasion games included, as it is the only one dependant on coordination between the eyes and the foot when executing.

The exclusion of studies measuring learning of so called especial skills (Breslin et al., 2012; Savelsbergh et al., 2012; Tillar & Marques, 2013), was with a backdrop of the differences between them and the rest of the studies. The goal of these studies was testing special situations in a stable environment, not like the invasion sports which are exerted in a unstable environment. Despite the exclusion, it has to be noted that they are not that dissimilar to some of the studies included. One of the studies including especial skills, (Breslin et al., 2012) measures learning of the ability to shoot free throws, and there are studies included using shooting from the free throw line as the testing measure of shooting ability in general (e.g. Memmert, 2006).

Looking back to the beginning of the process, the studies testing in a stable environment like the one mentioned from the free throw line in a self-paced tempo, may not be as close to the game context as needed when evaluating soccer learning. It may even have more in common with laboratory experiments, or other sports since there are no invasion areas in a stable environment. Anyhow, no matter the degree of similarity they were included, but the environment both in treatment and testing was used as moderating variables in the analysis.

When deciding on which tests to use in a study, or when evaluating the procedures like in this thesis, both the mentioned stability of environment and controllability are factors affecting application of tests. Game play is a test in an unstable environment, and

facilitates the similarity with the actual game context, increasing the ecological validity. The scientist gets the opportunity to see how the motoric skill is a part of the complex interactions in the central nervous system, which it arguably is in soccer (Ali et al., 2011). The downside being the lack of controllability, due to the subjective judgments of successful and non-successful actions (Ali, 2011).

When testing execution outcomes, the other option is skill tests, and when testing decision making dependable of tactical knowledge, perception and such, questionnaire (Gray & sproule, 2011) or video recognition tests using life size footage (i.g. Tallir et al., 2005) are other options. Contrary to game play, all these tests are more controllable, but in a stable environment.

The treatment groups were categorized initially by their degree of CI, compared to the other treatment groups in the same study. In the meta-analysis the effects with similar stability were clustered together, computing effects for unstable, stable and a combined stability (unstable group vs stable group) effect. Additionally, separating the effects by the amount of practice in an unstable and stable environment could have given more precise estimates, even more applicable for future practice. As the separation was based on the inclusion of at least partial practice in an unstable environment, the practice amounts in an unstable environment could potentially vary dramatically from study to study. The thesis includes comparisons of different amounts of unstable environment, but a comparison of for example 100% unstable practice, compared to 80% unstable practice stability could elaborate how to practice further.

Although the decision of solely including studies containing pre-test was made with a backdrop of recommendations, as the measure between pre to post observation benefits the measure of learning (Keppel, 1982), not all measures in the studies with significance findings were derived from testing more than post treatment (IGP design). This favours the use of IGPP design in the analysis. An even better measure would have been the use of multiple observations, as it makes it possible to draw growth curves (Keppel, 1982). The groups, when measuring with the IGP design, have to be based on a random selection of subjects if the selection effect is to be eliminated (Morris & Deshon, 2002).

A criticized criterion (Borenstein et al., 2009), is the use of impact factor. A similar way of evaluating studies is used in this thesis. The exclusion of studies published in journals not reported as level one or two (NSD, 2015). When excluding studies where the journals are stated as low quality, specifically, on the basis of reported level (NSD, 2015), its not entirely certain that the included studies are of high quality, and the excluded of low quality (Borenstein et al., 2009). NSD criteria for different levels are based on the guidelines made by Universitets og Høyskolerådet (2004), and as they state about their rating, it is hard to find suitable system which appropriately weights quality ahead of quantity, while also taking into account the distribution of money (see, Universitets og Høyskolerådet, 2004). With these problems, there are not a flawless system available at the present time (Universitets og Høyskolerådet, 2004), giving Borenstein and colleagues (2009) support for their recommendation of not using this as a basis of inclusion or exclusion. Although recommendations, the journals level of quality were evaluated (NSD, 2015) In retrospective manner, the levels of journals could have been chosen as a moderating variable instead, giving further information of the selection of studies, simultaneous increasing the quantity of studies.

A weakness of this thesis, as described in the general results, is the number of studies failing to state sufficient information of descriptive statistics. In some of the studies included the research report are imprecise. Characteristics missing in this thesis, like age and gender, should arguably be standard procedure of research, and they indeed are common variables overall (Weiner, Schinka & Velicer, 2012). In the case of systematic reviews, one of the purposes is to show where research is lacking (Dixon-Woods et al., 2006), and in these cases, when descriptive data are not shown, the lacking methodical qualities are shown. The same does not apply for the meta-analysis.

The missing data in this meta-analysis implied a further exclusion of some studies, like most meta-analysts would choose under similar circumstances (Cooper, 2010). There were several matters of incomplete information. One study (Turner & Martinek, 1999) was excluded, as the specified sample size of each treatment group was not given. Although not the only one, as the procedure used in the other instances was to divide the total sample size on the number of treatment groups, this was not possible in this incident. Specifically, the study had a total sample size of an odd number, making it impossible to estimate, as it included two treatment groups.

Two more studies were also excluded, although they included significant differences. The reason for this being insufficient information of raw scores, and simultaneous the significance results were given as F-ratios combining three groups in one study (Porter & Magill, 2010). If either, raw scores were stated, like in the quite similar study of Porter & Saemi (2010) where some of the outcomes included raw score, or a F-ratio score combining two groups was given, it could have been included (Morris & DeShon, 2002). As mentioned, not all outcomes in the study of Porter & Saemi (2010) had sufficient raw scores to be included. The last study to be excluded (Tallir et al., 2005), was excluded due to the use of ANCOVA in the statistical analysis, which is based on changed scores, and may not be included with the rest of the studies (Morris & DeShon, 2002).

4.3.2 Analysis/ results

Initially, in relation to the results, it has to be noted that in this thesis learning of the highest level of expertise possible was the purpose. The subjects in the studies have been assessed, solely on their performance at the tests, giving an estimate of learning in the treatment period. Many athletes do not have the urge to acquire the highest level of expertise (Malina, 2010), and for young athletes', lifelong engagement with the purpose of healthy living should be at least equally important. The conception that the generalization of results beyond the development of top level athletes, should not occur. If generalized across age and levels of expertise, independent of desire, it could be at the expense of values of higher order (Malina, 2010).

All estimates of, and in relation to, the effect was computed with raw scores. As the objective was to get as accurate conclusions as possible, it is important that all effect sizes uses the same type of data (Glass et al., 1981). The raw scores are recommended when testing different treatment methods, as stated in the literature about a similar study aim as mine, "*The effectiveness of a training program could be expressed as the difference between training and no training groups, suggesting raw scores metric*" (Morris & DeShon, 2002, p111). The other metric, changed score, would have the preferred one if measuring the size of the change as a result of training (Morris & DeShon, 2002). A positive side-effect of choosing the raw score metric, is that it does not get affected by the *p*-values enabling all measures from studies or samples to be

compared, in other words; there is no need for additional sub-groups with homogenous values (Morris & DeShon, 2002).

Hedges correction factor (1980) was applied, as described in the section of methods, with the objective of an unbiased estimator. Six effect sizes, deriving from four studies, were correct as they contained a sample size of 20 or less (Cheong et al., 2016; Landin & Hebert, 1997; Memmert, 2010; Tallir et al., 2007). Tallir and colleagues (2007) estimates had to be corrected using the IGP way described, see the section of methods. estimates had to be corrected using the IGP way described, see the section on methods. The backdrop being no available standard deviation from the groups separately. The missing data in the study, as well as a sample size of 20 or less could have been reasons for excluding the study, or at least careful use of it (Cooper, 2010). The reasoning for inclusion being the similar corrections no matter if the *d*-scores was estimated through a IGP or IGPP design.

A common standard deviation, when computing the effect sizes, assumes that the groups are similar (Cooper, 2010). If a pre-test standard deviation is used, as it is measured beforehand, it will properly produce more similar common standard deviation (Becker, 1988). This procedure is used when computing the effect sizes, when measures derive from IGPP design. Though, opposing to the use of a common pooled standard deviation some have argued that it will produce larger variability (Dunlap, Cortina, Vaslow & Burke, 1996), contradictive to the objective of standardization.

The outcomes from the IGP and IGPP designs was separated, both in the analysis and when presenting the results. When the deselection of combining the two designs could be justified both empirically, as the test of mean differences across designs showed systematic inequalities, and rationally, as the notion of time effect could not be ruled out based on the length of treatment periods across studies, the decision was made.

Its common to assume that the time effect between groups are equal, in relation to that changing is natural, but not when measuring in domains like training effect (Morris & DeShon, 2002). Both maturation (Malina et al., 2010) and previous experience can affect the effect of time, making it non zero (Morris & DeShon, 2002). The studies of this thesis selection reported up to 15 individual treatments, and 53% 8 treatments or

above. Therefore, the time between pre and post-test, as noted, may be an explanatory variable (Becker, 1988), and could have tested also in this thesis.

If these designs were aggregated, adjustments in the design would be necessary (Glass et al., 1981; Morris & DeShon, 1997), and estimation of the various sources of bias would be a part of the effects (Becker, 1988). In the research area of practice effectiveness, as in this thesis, it has been common that the pool of studies includes design with and without repetitive measures (Burke & Day, 1986). The separation of designs gives more precise and less biased measures, since also the pretest is evaluated (Morris & DeShon, 2002). The strengths by this choice facilitating the validity of the thesis, but at the same time weakening it, giving a lower quantity of comparable individual units of effects.

When presenting the effects, both when in the section of results and when discussing them later, measures from the IGPP design has been displayed when possible. As mentioned above, the use of two measures when measuring learning should not be underestimated, and if possible multiple observations should be applied, making it possible to observe the growth trajectory of subjects in comparison (Keppel, 1982). Though it has to be noted that all studies in this thesis included a pre-test, though not only when significant measures were found. In some cases, the pre-test was also used with the purpose of equal selection (e.g. Landing & Hebert, 1997) and to report non significant differences before the treatment (e.g. Porter & Saemi, 2010).

It is common practice to have a quantity threshold of individual units of analysis in meta analysis. In this thesis, three units were chosen (Harwood et al., 2015; Sallis, Prochaska & Taylor, 2000). Others have chosen different threshold, like Brady (2004) who had five units as a minimum, when presenting effects of CI in a broader motoric sense. The reasoning for the threshold, are the objective of getting meaningful summary (Harwood et al., 2015).

Comprehensive meta analysis (CMA), a computer program special made for meta analysis, was used in analysis after constructing the data needed. The use of computer programs simplifies the process when categorizing data (Mason, 2002), and may both strengthen and weaken the thesis. Mostly pros are associated with computer programs,

for instance, the fact that they are time saving (Thagaard, 2013) and strengthens the researchers' opportunity to creatively try alternate analysis, in the search of broader and deeper meaning (Corbin & Strauss, 2008), are pros to be considered. Both examples are important in this thesis. One con, which also applied in this thesis was the fact that, the process of categorizing data offered so many options, which occupied a lot of attention for a while (Mason, 2002).

CMA was chosen after searching the literature for appropriate programs to use. In a systematic comparison of computer programs dedicated to measure causality studies, through meta analysis, CMA scored highest both on usability and most complete sets of analytic features, when compared to five other programs. Another program scored equally, namely MIX (Bax, Yu, Ikeda & Moons, 2007).

The two most popular statistical models are the fixed-effects model, and the one chosen in this thesis, the random effects model. As chosen it, the effects in the studies represent a random selection of effects in the chosen domain. The goal being to measure the average distribution of effects (Borenstein et al., 2010). As noted earlier, fixed-effects model has only one source of variance, as it has only one source of sampling (Borenstein et al, 2010). An example could be a research conducted at the Norwegian school of science, solely including students in their first year which is similar across classes. They get randomly selected into groups. In this case so far there are only one source of variance, the students in the class. When applying the random effects model, the analysis takes into account more sources of variance (Borenstein et al., 2010). If modifying the example above, adding different schools, there will also be a variance dependable of schools. In this example, as it is possible that the effects also can vary in terms of schools, the fixed-effects model will no longer be appropriate.

In this thesis there are no reasons for assuming similarity across studies, relating to that the true effect are not completely identical. Even though the random-effects model takes into account more sources of variance, it does not explain it. Although some of the variance can be anticipated when studying moderator variables, and effects (Borenstein et al., 2010), for example age or stability in treatment in this study.

Analysing the *d*-effects, as mentioned, two methods of choosing individual units of analysis were used. The possibility of other options could also have been used, as the use of each researcher or laboratory, who is arguably the most conservative way. The negative with this option is that the researcher first has to compute a overall conclusion from each team of researchers or laboratory, conducting a synthesis before synthesizing between teams (Cooper, 2010). Though a decision of using a shifting unit of analysis was made, switching between using studies and samples as individual units of analysis.

When using individual units of samples (each significant outcome measure), the pro is that no information regarding potential moderator variables are lost (Cooper, 2010). A potentially important factor in this thesis, with the objective of differentiating research in soccer from general motoric research. The downside, is the fact that it can not be assumed a statistically independent estimate of each unit (Cooper, 2010). Studies as units, has the opposing pros and cons mentioned (Cooper, 2010).

When calculating several *d*-effects from the same observations, shifting the groups, the effects will no longer be statistically independent, as they are using the same average means and standard deviations. Nevertheless, it is the recommended procedure if including studies with more than two groups (Cooper, 2010). Other possible options are possible, but these are hard to use when conducting meta analysis, and therefore are rarely used (Cooper, 2010). For example, is it possible to calculate each groups percent of the variance from the dependable variable, though this method does not contain information about the treatment methods ranks and deviation (Cooper, 2010).

The *Q*-statistics was chosen as the test for homogeneity, but others could have been chosen. For example, using the Hunter & Schmidt (1990) 75% rule. *Q*-statistics could also have been computed for moderators, comparing them to the overall *Q*-statistics. This method is described by Hedges & Olkin (1985). Neither of these tests were computed, as limited time resulted in prioritizing. Since the original *Q* statistics can be computed both overall and on moderators this method was chosen.

After computing the effect of magnitude, in different variables, a statistical model of analysis could have been applied, with the objective of analysing the differences among groups means. For example, between the different age groups, youth and adult, testing

for significance like done in the analysis of Brady (2004). This was not applied as a matter of time consuming, but could have given more meaning, even though statistical significant should not be a substitute of scientific meaning (Wasserstein & Lazar, 2016).

As noted in the section of theoretical framework, the phenomenon in a population was considered as present or absent shown through statistical significant, before statistical power, and was missing a specific value on a common parameter (Brady, 2004). Statistical significant testing is arguably the most used approach when evaluating among social and behavioral scientists. When applied the major concern is that people misunderstand the true meaning of these results, but dependable on the interpretation it can be effective (Nickerson, 2000). No matter how interpretations are, it does not give a estimate of the strength in a relationship (Alderson & Green, 2008). Favouring thesis like this, who include and sum the magnitude through sizes of effect.

5. Conclusion

A total of 17 studies was included, and 12 reported significant differences when comparing practice schedules with high and low degree of CI. Overall, a large learning effect was found favouring high degrees of CI, supporting the original notion (Battig, 1966) that a random and variable practice schedule is superior on the effect of learning. This effect was the general direction, and seems robust when considering the significant homogeneity test. One more effect had a significant homogeneity test, the outcome decision making. These two effects are the most robust.

More effects of different sizes were found, but no other with the same homogeneity. Although methodical implication many shows trends, and may be used as implications. Soccer as a specific domain, seems to exceed the general learning effect found in broader motor domains (e.g. Brady, 2004). Although Soccer was the primary sport, the effect can be seen as large across sports, as the thesis included several different invasion sports. For a better sense of an even more meaningful effect for Soccer more studies are needed, as only four studies were included in this thesis.

As for treatments differing in stability of environment, an unstable one and a higher degree of CI seems superior. The only exception seems to be when practicing in a stable environment, where a moderate degree of CI seems most beneficial for learning. When moderating for testing procedures, game play had a large effect and the skill test only a small effect. Implying a larger effect, in a more similar environment the competition context, but game play has the weaknesses in lack of controllability. The earlier shown differences when testing for age, test time point did not seem to differ. Also here the high degree of CI seems superior.

Future research should increase the specificity of the studies. A step in the right direction is the study of Cheong and colleagues (2016), which include three groups, different stability in the treatment environment, and both gameplay and skill tests. This is also the article introducing the stability of environment notion, which should be an implication for future research.

6 References

- Abrantes, C. I., Nunes, M. I., Macas, V. M., Leite, N. M. Sampaio, J. E. (2012). Effects of the number of players and game type constraints on heart rate, rating of perceived exertion, and technical actions of small-sided soccer games. *Journal of Strength and Conditioning Research*, 26 (4), p976-981.
- Alderson, P & Green, S. (2008). *Cochrane collaboration open learning materials for reviewers*. The Cochrane Collaboration.
- Ali, A. (2011). Measuring soccer skill performance: a review. *Scand J Med Sci Sports*, 21, p170-183.
- Ali, A., Foskett, F. & Gant, N. (2008). Validation of a soccer skill test for use with female players. *Int J Sports Med*, 29, p191-197.
- Ali, A., Williams, C., Hulse, M., Strudwick, A., Reddin, J., Howarth, L., Eldred, J. E., Hirst, M. & McGregor, S. J. (2007). Reliability and validity of two tests of soccer skill. *J Sports Sci*, 25, p1461-1470.
- Allison, S. & Thorpe, R. (1997). A comparison of the effectiveness of two approaches to teaching games within physical education. A skills approach versus a games for understanding approach. *The British Journal of Physical Education*, 28 (3), p9-13.

- Arponen, V. P. J. (2013). The extent of cognitivism. *History of the Human Sciences*, 26 (5), p3-21.
- Baker, J. (2007). Nature and nurture interact to create expert performers. *High ability studies*, 18 (1), p57-58.
- Baker, J., Cote, J., Abernethy, B. (2010). Sport-specific practice and the development of expert decision-making in team ball sports. *Journal of Applied Sport Psychology*, 15 (1), p12-25.
- Barnsley, R. H., Thompson, A. H. & Legault, P. (1992). Family planning: football style. The relative age effect. *International Review for the Sociology of sport*, 27 (1), p77-87.
- Barreiros, J., Figueiredo, T. & Godinho, M. (2007). The contextual effect in applied settings. *European physical education review*, 13 (2), p195-208.
- Bate, D. (1996). Soccer skills practice. In: T. Reilly (Red.). *Science and soccer*, (p227-241). London: Routledge.

Battig, W. E. (1966). Facilitation and interference. In: E. A. Bilodeau (Ed.). *Acquisition of skill*, (p215-244). New York: Academic Press.

Bax, L., Yu, L., Ikeda, N. & Moons, K. G. M. (2007). A systematic comparison of software dedicated to meta analysis of causal studies. *BMC Medical Research Methodology*, 7 (40), p1-9.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical & Statistical Psychology*, 41 (2), p257-278.

Berry, J., Abernethy, B., Cote, J. (2008). The contribution of structured activity and deliberate play to the development of expert perceptual and decision-making skill. *Journal of Sport & Exercise Psychology*, 30, p685-708.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. UK: Wiley

Borenstein, M., Hedges, L. V., Higgins, J. P. T & Rothstein, H. R. (2010). A basic Introduction to fixed and random-effects models for meta-analysis. *Research Synthesis Methods*, 1, p97-111.

- Bunker, D. & Thorpe, R. (1982). Model for the of games in secondary schools. *Bulletin of Physical Education Spring, 18*, p5-8.
- Burke, M. J. & Day, R. R. (1986). A cumulative study of effectiveness of managerial training. *Journal of Applied Psychology, 71 (2)*, p232-245.
- Butler, J. (2014). TGfU – Would you know it if you saw it? Benchmarks from the tacit knowledge of the founders. *European Physical Education Review, 20 (4)*, p465-488.
- Brady, F. (1998). A theoretical and empirical review of the contextual interference effect and the learning of motor skills. *Quest, 50*, p266-293.
- Brady, F. (2004). Contextual interference: a meta-analytic study. *Perceptual and Motor Skills, 99*, p116-126.
- Brady, F. (2008). The contextual interference effect and sport skills. *Perceptual and motor skills, 106*, p461-472.
- Breslin, G., Hodges, N. J., Steenson, A. & Wiliams, A. M. (2012). Constant or variable practice: Recreating the especial effect. *Acta Psychologica, 140*, p154-157.

- Broadbent, D. P., Causer, J., Williams, A. M. & Ford, P. R. (2015). Perceptual-cognitive skill training and its transfer to expert performance in the field: Future research directions. *European Journal of Sport Science*, 15 (4), p322-331.
- Brobst, B. & Ward, P. (2002). Effects of public posting, goal setting, and oral feedback on the skills of female soccer players. *Journal of Applied Behavior Analysis*, 35 (3), p247-257.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.) Berkley, USA: University of California Press.
- Causer, J. & Ford, P. R. (2014). "Decisions, decisions, decisions": Transfer and specificity of decision-making skill between sports. *Cognitive Processing*, 15 (3), p385-389.
- Campbell, D. T. & Stanley, J. C. (1966). *Experimental and quasi-experimental design for research*. Chicago: Rand McNally.
- Chatzopoulos, D., Drakou, A., Kotzamandidou, M. & Tsorbatzoudis, H. (2006a). Girls's soccer performance and motivation: games vs technique approach. *Perceptual and Motor Skills*, 103, p463-470.

- Chatzopoulos, D., Tsormbatzoudis, H. & Drakou, A. (2006b). Combinations of technique and games approaches: Effects on game performance and motivation. *Journal of Human Movement Studies*, 50, p157-170.
- Cheong, J. & Lay, B. (2013). Is the contextual interference effect supported when practicing several skills in combination?. *Journal of Science and Medicine in Sport*, 16 (1), p1-72.
- Cheong, J. P. G., Lay, B., Grove, J. R., Medic, N. & Rasman, R. (2012). Practicing field hockey skills along the contextual interference continuum: A comparison of five practice schedules. *Journal of Sports Sciences and Medicine*, 11, p304-311.
- Cheong, J. P. G., Lay, B. & Razman, R. (2016). Investigating the contextual interference effect using combination sports skills in open and closed skill environments, *Journal of Sports Science and Medicine*, 15, p167-175.
- Cheong, J., Lay, B., Grove, R. & Razman, R. (2010). The effect of block, random and combination practice order on learning variations of the same task. *Journal of science and Medicine in Sport*, 13 (1), p1-42.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), p155-159.

Comprehensive Meta-analysis (3.0). [CMA software]. Retrieved 24 Marc from:
https://www.meta-analysis.com/pages/why_use.php

Conte, L., Moreno-Murcia, J. A., Perez, G. & Iglesias, D. (2013). Comparison of traditional and understanding methodology in the practice of basketball. *Revista Internacional de Medicina y Ciencias de la Actividad Fisica y el Deporte*, 13 (51), p507-523.

Copper, H. (2010) *Research Synthesis and Meta-Analysis; A Step by Step Approach* (4th ed.). Thousand Oaks, CA:Sage.

Corbin, J. M. & Strauss, A. (2008). *Basics of qualitative research: techniques and procedures for developing grounded theory (3rd ed.)*. London: Sage.

Davids, K., Araujo, D. Hritovski, R., Passos, P. & Chow, J. J. (2012). Ecological dynamics and motor learning design in sport. In: N. J. Hodges & A. M Williams. *Skill acquisition in sport – research, theory and practice (2th ed.)*, (p112-130). London and New York: Routledge.

- Davids, K. & Baker, J. (2007). Genes, environment and sport performance- why the nature-nurture dualism is no longer relevant. *Review Article Genes, Environment and Sport Performance*, 37 (11), p961-980.
- Davids, K., Williams, A. M., Button, C. & Court, M. (2001). An integrative modeling approach to the study of intentional movement behavior. I: R. N. Singer, H. A. Hausenblas & C. M. Janelle (Ed.), *Handbook of sport psychology (2th ed.)*, (p144-173). New York: Wiley.
- Deprez, D. N., Franssen, J., Lenoir, M., Philippaerts, R. M. & Vaeyens, R. (2015). A retrospective study on anthropometrical, physical fitness and motor coordination characteristics that influence dropout, contract status, and first-team playing time in high-level soccer players aged eight to eighteen years. *Journal of Strength & Conditioning Research*, 29 (6), p1692-1705.
- Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, Hsu, R., Katbamna, S. Olsen, R., Smith, L., Roley, R. & Sutton, A. J. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research methodology*, 6 (35), p1-13.
- Duncan, C. P. (1953). Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *Journal of experimental psychology*, 45 (1), p1-11.

- Dunlap, W. P., Cortina, J. M., Vaslow, J. B. & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological methods, 1* (2), p170-177.
- Elferink-Gemser, M. T., Visscher, C., Lemmink, K A. P. M. & Mulder, T. W. (2004). Relation between multidimensional performance characteristics and level of performance in talented youth field hockey players. *Journal of Sports Sciences, 22*, p1053-1063.
- Ericsson, K. A. (1996). *The Road to excellence; The acquisition of expert performance in arts and sciences, sports and games*. Hillsdale, NJ: Erlbaum.
- Ericsson, K. A., Krampe, R. T. & Tesch- Roemer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, p363-406.
- Ericsson, K. A. & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, p273-305.
- Ericsson, K. A. & Smith, J. (1991). *Toward a grounded theory of expertise: Prospects and Limits*. Cambridge: Cambridge University Press.

- FIFA (2007). FIFA Big count 2006. *Report from FIFA Communication Divisions*.
Retrieved 23 of May 2016 from http://www.fifa.com/mm/document/fifafacts/bcoffsurv/bigcount.statspackage_7024.pdf
- Figueiredo, A. J., Goncalves, C. E., Coelho e Silva, M. J. & Malina, R. M. (2009).
Characteristics of youth soccer players who drop out, persist or move up.
Journal of Sports Sciences, 27 (9), p883-891.
- Figueiredo, A. J. & Coelho e Silva, M. J. (2010). Size and maturity mismatch in youth
soccer players 11- to 14-years-old. *Pediatric Exercise Science*, 22, p596-612.
- Ford, P. R., Yates, I. & Williams, A. M. (2010). An analysis of practice activities and
instructional behaviors used by youth soccer coaches during practice: exploring
the link between science and application. *Journal of Sports Sciences*, 28 (5),
p483-495.
- French, K. E., Werner, P. H., Rink, J. E., Taylor, K. & Hussey, K. (1996). The effects of
a 3-week unit of tactical, skill or combined tactical and skill instruction on
badminton performance of ninth-grade students. *Journal of Teaching in Physical
Education*, 15 (4), p418-438.
- Glass, G. V., McGaw, B. & Smith, M. L. (1981). *Meta-analysis in social research*.
Beverly Hills, CA: Sage.

- Gibson, J. J. & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment?. *Psychological Review*, 62 (1), p32-41.
- Giulianotti, R. & Robertson, R. (2004). The globalization of football: A study in the glocalization of the 'serious life'. *The British Journal of Sociology*, 55, p545-568.
- Goodger, K., Gorely, T., Lavalee, D. & Harwood, C. (2007). Burnout in sport: A systematical review. *The Sport Psychologist*, 21, p127-151.
- Gray, S. & Sproule, J. (2011). Developing pupils' performance in team invasion games. *Physical Education and Sport Pedagogy*, 16(1), p15-32.
- Grehaigne, J., Godbout, P. & Bouthier, D. (2001). The teaching and learning of decision making in team sports. *Quest*, 53 (1), p59-76.
- Griffin, L. L., Brooker, R. & Patton, K. (2005). Working towards legitimacy: Two decades of teaching games for understanding. *Physical Education and Sport Pedagogy*, 10 (3), p213-223.
- Griffin, L. L. & Butler, J. I. (2005). *Teaching Games for Understanding*. USA: Human Kinetics.

- Griffin, L. L., Mitchell, S. A. & Oslin, J. L. (1997). *Teaching sports concepts and skills: A tactical games approach*. Champaign, IL: Human Kinetics.
- Guadagnoli, M. A., Holcomb, W. R. & Weber, T. J. (1999). The relationship between contextual effects and performer expertise on the learning of a putting task. *Journal of Human Movement Studies*, 37, p19-36.
- Guadagnoli, M. A. & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, 36, p212-224.
- Harwood, C. G., Keegan, R. J., Smith, J. M. J. & Raine, A. R. (2015). A systematical review of the intrapersonal correlates of motivational climate perceptions in sport and physical activity. *Psychology of Sport and Exercise*, 18, p9-25.
- Hassmen, N. & Hassmen, P. (2008). *Idrottsvetenskapliga forskningsmetoder*. Sweden: SISU Idrottsböcker
- Haugaasen, M., Toering, T., Jordet, G. (2014). From childhood to senior professional football: A multi-level approach to elite youth football players engagement in football-specific activities. *Psychology of Sport and Exercise*, 15, p336-344.

Hedges, L. V. (1980). Unbiased estimation of effect size. *Evaluation in Education*, 4, p25-27.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic press.

Helsen, W. F., Van Winckel, J. V. & Williams, A. M. (2005). The relative age effect in youth soccer across europe. *Journal of sports sciences*, 23 (6), p629-636.

Higgins, J. P. T. & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions*. The Cochrane collaboration. <http://handbook.cochrane.org/>

Holt, J. E., Ward, P. & Wallhead, T. L. (2006). The transfer of learning from play practices to game play in young adult soccer players, *Physical Education and Sport Pedagogy*, 11 (2), p101-108.

Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta analysis: Correcting error and bias in research findings*. California: Sage

Huijgen, B. C. H., Elferink-Gemser, M. T., Lemmink, K. A. P. M & Visscher, C. (2014). Multidimensional performance characteristics in selected and deselected soccer players. *European Journal of Sport Sciences*, 14 (1), p2-10.

- Janelle, C. M. & Hillman, C. H. (2003). Expert performance in sport: Current perspectives and critical issues. In: J. L. Starkes & K. A. (Ed.), *Expert performance in sport: Advances in research on sport expertise* (p19-48). Champaign, IL: Human Kinetics.
- Kannekens, R., Elferink-Gemser, M, T. & Visscher, C. (2009). Tactical skills of world-class youth soccer teams. *Journal of Sports Sciences*, 27 (8), p807-812.
- Keppel, G. (1982). *Design and analysis: a researcher's handbook* (2nd ed.). New Jersey: Prentice Hall.
- Knapp, B. (1977). *Skill in sport: The attainment of proficiency*. London: Routledge.
- Landin, D. & Hebert, E. P. (1997). A comparison of three practice schedule along the contextual interference continuum. *Research Quarterly for Exercise and Sport*, 68 (4), p357-361.
- Landin, D. K., Hebert, E. P. & Fairweather, M. (1993). The effects of variable practice on the performance of a basketball skill. *Research Quarterly for exercise and Sport*, 64 (2), p232-237.

- Landin, D., Hebert, E. P. Menickelli, J. & Grisham, W. (2003). The contextual interference continuum: What level of interference is best for adult novices?. *Journal of Human Movement Studies*, 44, p19-35.
- Lauder A. G. (2001). *Play practice: The games approach to teaching and coaching sports*. Champaign, IL: Human Kinetics
- Le Gall, F., Carling, C., Williams, M. & Reilly, T. (2010). Anthropometric and fitness characteristics of international, professional and amateur male graduate players from an elite youth academy. *Journal of Science and Medicine in Sport*, 13 (1), p90-95.
- Lee, T. D. & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In: D. Goodman, R. B. Wilberg & I. M. Franks (Red.), *Differing perspectives in motor learning, memory, and control*, (p3-22). Amsterdam: Elsevier.
- Li, Y. & Lima, R. P. (2002). Rehearsal of task variations and contextual interference effect in a field setting. *Perceptual and Motor Skills*, 94, p750-752.
- Lin, C- J., Wu, A. D., Udompholkul, P. & Knowlton, B. J. (2010). Contextual Interference effects in sequence learning for young and older adults. *Psychology and Aging*, 25 (4), p929-939.

- Magill, R. A. (1992). Practice schedule considerations for enhancing human performance in sport. In: R. Christina & H. Eckert (Ed.). *Enhancing human performance in sport: New concepts and developments*, (p38-50). IL: Human Kinetics.
- Magill, R. A. & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science*, 9, p241-289.
- Malina, R. M., Pena Reyes, M. E., Eisenmann, J. C. Horta, L., Rodrigues, J. & Miller, R. (2000). Height, mass, and skeletal maturity of elite Portuguese soccer players aged 11-16 years. *Journal of Sport Sciences*, 18, p683-693.
- Malina, R. M. (2010). Early sport specialization: roots, effectiveness, risks. *Current Sports Medicine Reports*, 9, p364-371.
- Malina, R. M., Pena Reyes, M. E., Figueiredo, A. J., Coelho E Silva M. J., Horta, L., Miller, R. & Morate, F. (2010). Skeletal age in youth soccer players: Implication for age verification. *Clin J Sport Med*, 20 (6), p469-474.
- Mason, J. (2002). *Qualitative Researching (2nd ed.)*. London: Sage.
- McMorris, T. & Graydon, J. (1996). The effect of exercise on the decision-making performance of experienced and inexperienced soccer players. *Research Quarterly for Exercise and Sport*, 67 (1), p109-114.

- McMorris, T. & Graydon, J. (1997). The effect of exercise on cognitive performance in soccer-specific tests. *Journal of Sports Sciences*, 15, p459-468.
- Memmert, D. (2006). Long-term effects of type of practice on the learning and transfer of a complex motor skill. *Perceptual and Motor Skills*, 103, p912-916.
- Memmert & Roth. (2006). The effects of non-specific and specific concepts on tactical creativity in team ball sports. *Journal of Sport Sciences*, 25 (12), p1423-1432.
- Merbah, S. & Meulemans, T. (2011). Learning a motor skill: effects of blocked versus random practice a review. *Psychologica Belgica*, 51 (1), p15-48.
- Mesquita, I., Farias, C. & Hastie, P. (2012). The impact of a hybrid sport education-invasion games competence model soccer unit on students decision making, skill execution and overall game performance. *European Physical Education review*, 18 (2), p205-219.
- Mitchell, S. A. & Oslin, J. L. (1999). An investigation of tactical transfer in net games. *European Journal of Physical Education*, 4, p162-172.
- Mohr, M., Krstrup, P & Bangsbo, J. (2003). Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sport Sciences*, 21 (7), p519-529.

- Morris, S. B. & DeShon, R. P. (1997). Correcting effect sizes computed from factorial analysis of variance in meta-analysis. *Psychological methods*, 2 (2), p192-199.
- Morris, S. B. & DeShon, R. P. (2002). Combining effect size estimates in meta- analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7 (1), p105-125.
- Morris, T. (2000). Psychological characteristics and talent identification in soccer. *Journal of Sport Sciences*, 18 (9), p715-726.
- Murtagh, M. J., Thomson, R. G., May, C. R., Rapley, T., Heaven, B. R., Graham, R. H., Kaner, E. F., Stobbart, L. & Eccles, M. P. (2007). Qualitative methods in a randomized controlled trial: the role of an integrated qualitative process evaluation in providing evidence to discontinue the intervention in one arm of a trial of a decision support toll. *Quality and Safety in Health Care*, 16, p224-229.
- Musch, J. & Hay, R. (1999). The relative age effect in soccer: cross-cultural evidence for a systematic discrimination against children born late in the competition year. *Sociology of Sport Journal*, 16, p54-64.
- Newell, K. M. & McDonald, P. V. (1992). Practice: A search for task solutions. *American Academy of Physical Education Papers*, 25, p51-59.

- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological methods*, 5 (2), p241-301.
- NSD (2015). Publiseringkanaler. *Database for statistikk om høgre utdanning*. Retrieved 27 May 2016 from: <https://dbh.nsd.uib.no/publiseringkanaler/Forside>
- Olosova, G. & Zapletalova. (2015). Immediate and retention effects of teaching games for understanding approach on basketball knowledge. *Acta Facultatis Educationis Physicae Universitatis Comenianae*, 55 (1), p39-45.
- Pease, D. G. & Pupnow, A. A. (1983). Effects of varying force production in practice schedule of children learning a discrete motor task. *Perceptual and Motor Skills*, 37, p275-282.
- Porter, J. M. & Magill, R. A. (2010). Systematical increasing contextual interference is beneficial for learning sport skills. *Journal of Sports Sciences*, 28 (12), s1277-1285.
- Porter, J. & Saemi, E. (2010). Moderately skilled learners benefit by practicing with systematic increase in contextual interference. *International Journal of Coaching Science*, 4 (2), p61-71.

- Priklerová, S. & Kucharik, I. (2015). Efficiency of technical and tactical approach to *Educationis Physicae Universitatis Comenianae*, 55 (2), s132-140.
- Psotta, R. & Martin, A. (2011). Changes in decision making skill and execution in soccer performance: The intervention study. *Acta Uni. Palacki. Olomuc., Gymn.*, 41 (2), p7-15.
- UEFA (2015). *UEFA Where the money goes*. Champions League revenue distribution. Retrieved 27 May from:
<http://www.uefa.com/uefachampionsleague/news/newsid=2294532.html>
- Reilly, T. & Gilbourne, D. (2003). Science and football: a review of applied research in the football codes. *Journal of Sports Sciences*, 21 (9), p693-705.
- Reilly, T. Williams, A. M., Nevill, A. & Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sport Sciences*, 18, p695-702.
- Roen, K., Arai, L., Roberts, H. & Popay, J. (2006). Extending systematic reviews to include evidence on implementation: methodological work on a review of community-based initiatives to prevent injuries. *Social Science in Medicine*, 63, p1060-1071.

- Ronglan, L. T. (2008). *Lagspill, læring og ledelse – om lagspillemens didaktikk*. Norge: Akilles.
- Rösch, D. Hodgson, R., Peterson, I., Graf-Baumann, T. Junge, A., Chomiak, J, Dvorak, J. (2000). Assessment and evaluation of football performance. *Am J Sports Med*, 28, p29-39.
- Rosenthal, R. (1978). How often are our numbers wrong? *American psychologist*, 33 (11), p1005-1008.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86 (3), p638-641.
- Rushall, B. S. & Pyke, F. S. (1990). *Training for Sport and Fitness*. Sydney: Macmillan.
- Sallis, J. F., Prochaska, J. J. & Taylor, W. C. (2000). A review of correlates of physical activity of physical activity of children and adolescents. *Medicine & Science in Sport & Exercise*, 32 (5), p963-975.
- Savelsbergh, G., Canal-Bruland R. & Kamp, J. V. D. (2012). Error reduction during practice: a novel method for learning to kick free-kicks in soccer. *International Journal of Sports Science & Coaching*, 7 (1), p47-56.

Schmidt, R. A. (1975). A Schema theory of discrete motor skill learning. *Psychological Review*, 82, p225-260.

Schmidt, R. A. & Lee, T. D. (2011). *Motor control and learning: A behavioral emphasis (5th ed.)*. Champaign, IL: Human Kinetics.

Seale, C. (2003). Quality in qualitative research. In: Y. S. Lincoln & Denzin. *Turning points in qualitative research – tying knots in a handkerchief* (p143-168). USA: AltaMira Press.

Shadish, W. R. & Haddock, C. K. (1994). Combining estimates of effect size. Cooper, H. & Hedges, L. V. (Ed.), *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

Shapiro, D. C. & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental limitations. In: J. A. S. Kelso & J. E Clark (Red.), *The development of movement control and coordination* (p113-150). New York: Wiley.

Shea, J. B. & Morgan, R, L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, 5, p179-187.

- Shea, J. B. & Zimny, S. T. (1983). Context effects in memory and learning information. In: R. A. Magill (Ed.), *Memory and control of action* (p345-366). Amsterdam: North Holland.
- Silverman, D. (2011). *Interpreting qualitative data: a guide to the principles of qualitative research (4th ed.)*. London: Sage.
- Silva, P., Duarte, R., Sampaio, J., Aguiar, P., Davids, K., Araujo, D. & Garganta, J. (2014). Field dimensions and skill level constrain team tactical behaviours in small sided and conditional games in football. *Journal of Sports Sciences*, 32 (20), p1888-1896.
- Smeeton, N. J., Ward, P., Williams, A. M. (2004). Do pattern recognition skills transfer across sports? A preliminary analysis. *Journal of Sports Sciences*, 22, p205-213.
- Stock, W. A., Okun, M. A., Haring, M. J., Miller, W. Kinney, C. & Ceuvorst, R. W. (1982). Rigor in data synthesis: case study of reliability in meta analysis. *Educational researcher*, 11 (6), p10-14+20.
- Stratton, G., Reilly, T., Richardson, D. & Williams, A. M. (2004). *Youth soccer: from science to performance*. London: Routledge

Tallir, I. B., Lenoir, M., Valcke, M., Musch, E. (2007). Do alternative instructional approaches result in different game performance learning outcomes? Authentic assessment in varying game conditions. *International Journal of Sports Psychology*, 38 (3), s263-282.

Tallir, I. B., Musch, E., Valcke, M. & Lenoir, M. (2005). Effects of two instructional approaches for basketball on decision-making and recognition ability. *International Journal of Sport Psychology*, 36 (2), p107-126.

Thagaard, T. (2013). *Systematikk og innlevelse – en innføring i kvalitativ metode*, (4th ed.). Bergen: Fagbokforlaget.

Thorpe, R, Bunker, D & Almond, L. (1986). *Rethinking games teaching*. UK: Loughborough.

Tillaar, R. V. D. & Marques, M. C. (2013). Effect of specific versus variable practice upon overhead throwing speed in children. *Perceptual & Motor Skills: Physical Development & Measurement*, 116 (3), p872-884.

Toering, T. T., Elfering-Gemser, M. T., Jordet, G. & Visscher, C. (2009). Self-regulation and performance level of elite and non-elite youth soccer players. *Journals of Sports Sciences*, 27 (14), p1509-1517.

- Turner, A. P. & Martinek, T. J. (1992). A comparative analysis of two models for teaching games (technique approach and game-Centered (tactical focus) approach). *International Journal of Physical Education*, 29 (4), p15-31.
- Turner, A. P. & Martinek, T. J. (1999). An investigation into teaching games for understanding effects on skill, knowledge, and game play. *Research Quarterly for Exercise and Sport*, 70(3), p286-296.
- Universitets- og høyskolerådet (2004). Vekt på forskning – nytt system for dokumentasjon av vitenskapelig publisering. *Instilling fra faglig og teknisk utvalg til HLR*. Retrived 27 may from:
http://www.uhr.no/documents/Vekt_p_forskning_sluttrapport.pdf
- Van Rossum, J. H. A. (1990). Schmidts schema theory: the empirical base of the variability of practice hypothesis. *Human Movement Science*, 9, p387-435.
- Vera, J. G., Alvarez, J. C. B. & Medina, M. M. (2008). Effects of different practice conditions on acquisition, retention, and transfer of soccer skills by 9-year old schoolchildren. *Perceptual and Motor skills*, 106, p447-460.
- Ward, P., Hodges, N. J., Williams, A. M. & Starkes, J. L. (2004). Deliberate practice and expert performance: Defining the path to excellence. In A. M. Williams and N. J. Hodges (Ed.), *Skill acquisition in sport: Research, theory and practice* (p231-258). London Routledge.

- Ward, P., Hodges, N. J., Starkes, J. L. & Williams, M. A. (2007). The road to excellence: Deliberate practice and development of expertise. *High Ability Studies, 18*, p119-153.
- Wasserstein, R. L. & Lazar, N. (2016). The ASA's statement on p-values: context, process, and purpose. *The American statistician*. DOI: 10.1080/00031305.2016.1154108
- Weiner, I. B., Schinka, J. A. & Velicer, W. F. (2012). *Handbook of psychology – volume 2 - researching methods in psychology* (2nd ed.). London: Wiley
- Williams, A. M. (2000). Perceptual skill in soccer: Implications for talent identification and development. *Journal of Sports Sciences, 18 (9)*, p737-750.
- Williams, A. M., & Hodges, N. J. (2005). Practice, instruction and skill acquisition: Challenging tradition. *Journal of Sports Sciences, 23 (6)*, p637-650.
- Williams, A. M. & Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sport Sciences, 18 (9)*, p657-667.
- Wright, S., McNeill, M., Fry, J. & Wang, J. (2005). Teaching teachers to play and teach games. *Physical Education and Sport Pedagogy, 10 (1)*, p61-82.

- Wrisberg, C. A. & Mead, B. J. (1981). Anticipation of coincidence in children: A test of schema theory. *Perceptual and Motor skills*, 52, p599-606.
- Wulf, G. & Schmidt, R. A. (1994). Feedback induced variability and the learning of generalized motor programs. *Journal of Motor Behavior*, 26, p348-361.
- Wulf, G., Shea, C. & Lewthwaite, R. (2010). Motor skill learning and performance: a review of influential factors. *Medicine Education*, 44, p75-84.
- Yan, J. H., Thomas, J. R. & Thomas, K. T. (1998). Children's age moderates the effect of practice variability: A Quantitative review. *Research Quarterly for Exercise and Sport*, 69 (2), p210-215.
- Young, D. E., Cohen, M. J. & Husak, W. S. (1993). Contextual interference & motor skill acquisition: one the processes that influence retention. *Human Movement Science*, 12, p577-600.
- Zuffová, Z. & Zapletalová, L. (2015). Efficiency of different teaching models in teaching frisbee ultimate. *Acta Facultatis Educationis Physicae Universitatis Comenianae*, 55 (1), p64- 73.

7. List of Figures

Figure 1: An in-detail overview of the search computed, including key word combinations, total number of studies found, and the final studies included in the review..... 25

Figure 2: An overview of the inclusion and exclusion process. 26

8. List of Tables

Table 1: Included studies, categorized by type of practice schedule and degree of contextual interference.....	28
Table 2: A descriptive overview of study and subject characteristics.	35
Table 3: Describes the key features relevant in the treatment and testing procedure, who are common across groups, as well as significant differences in learning between groups.....	36
Table 4: Stability of environment in treatments of high and moderate degree of contextual interference in studies with three groups. Quantity, outcome and direction of the significant differences between groups.....	38
Table 5: An overview of moderating variables effects, and their estimates (quantity of individual effect sizes, standard error, confidence interval, classification of magnitude and homogeneity).....	40
Table 6: Overview of the execution outcomes specifics, with their effects, and their estimates (quantity of individual effect sizes, standard error, confidence interval, classification of magnitude and homogeneity).	41
Table 7: Effects of different stabilities in case of the environment in the treatment and testing procedures,, and their estimates (quantity of individual effect sizes, standard error, confidence interval, classification of magnitude and homogeneity).....	42

9. Abbreviations

CI	Contextual Interference
CMA	Comprehensive Meta-Analysis
DPF	Deliberate practice framework
GBA	Game Based Approach
IGP	Independent Groups Post design
IGPP	Independent Groups Pre to Post design
MIX	Meta Analysis in Excel
NSD	Traditional Approach
PP	Play practice
RAE	Relative Age Effect
TA	Traditional Approach
TGFU	Teaching Games for Understanding

