

Does 360-degree Feedback work?

An exploratory study about the effect of 360-degree feedback assessments over time.



214233

Borsten, M.C.J.

Turnhoutplantsoen 108

1066 NW Amsterdam

Supervisor Tilburg University: Brigitte Kroon

Research guidance: Jan Tisauwer

Project Period: Spring semester 2010

Project Theme: 360 degree or multi-source feedback/ Keiwijzer

Introduction

One of the best ways of getting to know ourselves is by asking others. According to the Johari Window (Luft & Ingham, 1955) every individual has some 'blind spots'. These are behavioural characteristics that an individual is not aware of, but are recognized by others, most of the time by those who are closely related to the individual. Receiving feedback about oneself can shed light on one's performance. There are many different types of feedback, but in this study the focus is on 360 degree (or multi-source) feedback, which is especially good in paying attention to this issue of the 'blind spots'.

The reason why 360 degree feedback assessments are specialized in looking for 'blind spots' is because, with this form of feedback, performance evaluations are collected from constituencies representing the full circle of relevant raters, rather than collecting the feedback from only one rater. Most of the time the raters are subordinates, colleagues and supervisors, but also taken into the equation are sometimes other raters, like customers or assessors (Warech, Smither, Reilly, Milsap & Reilly, 1998).

In 1994, Brian O'Reilly wrote an article for Fortune magazine in which he addressed the importance of 360 degree feedback. In several interviews with key executives of major American companies they concluded that 360 degree feedback had an enormous *positive* impact on the professional career of their employees. This emphasizes the importance of feedback in the professional business world.

After this brief introduction of what 360 degree feedback is and the need for good feedback is underlined, the interest for this study shifts towards the period before and after the assessments, rather than during the assessment. This is because, in this study the interest is not in whether an organization should or should not use feedback or how people should fill in the assessment, but in how to make it as effective and efficient as possible, both in preparation and aftercare of the assessment.

First, we will discuss the four stages of Ilgen, Fisher and Taylor (1979), which are concerned with the period after the 360 degree feedback assessment. Ilgen et al. (1979) were pioneers in the field of 360 degree feedback assessment and therefore several authors have criticized their structuring of the stages over the years (Greenhaus, Callanan and Godshalk, 2009). Nevertheless, their foundation is still relevant and considered an added value for this study, because a lot of recent articles still draw upon their work (Shute, 2008; Goodman & Wood, 2009).

Before discussing each stage extensively in the theoretical framework, each stage will be introduced here shortly. The first stage is the perception stage; the second stage is called the

acceptation stage; the third stage is the stage where the individual decides on appropriateness for future development; and, in the fourth and final stage, behavioural change will commence. After someone has filled in his or her 360 degree feedback assessment and received the results from his peers, it is possible to write an individual development plan based on their results. But before this happens each stage must be encountered, which starts with the first stage called "Perception of feedback" (Ilgen et al., 1979). Before the individual development can start the results need to be interpreted by the individual (ratee), because the ratee needs to know what their shortcomings are in order to decide what to improve.

These stages will be discussed because this research will try to explore the great number of influences, such as ratees' emotions, assessment preparation and future guidance that need to be understood in order to make sure the 360 degree feedback assessment is able to establish a performance improvement. Without each stage being under control by the assessment giver, it is very likely that in each stage there is the possibility of a waste of energy of ratee, raters and investor (most likely the company the ratee works for).

Two of the above introduced influences have already been discussed. Assessment preparation, however, is not. For this we turn to the work of Fleenor, Taylor and Chappelow (2008). In their research they introduced some basic premises for successful 360 degree feedback assessments. Some of these premises will be introduced here to give a first impression their contents. Fleenor et al. (2008) describe that 360 degree feedback process works best if it begins with the top executives of the organization and then cascades through the organization. Another premise states that an untidy administration of a 360 degree feedback process can be fatal to future administrations. A third one notes the need for careful selection process in choosing and selecting raters. The ratee might decide to hand out assessment forms to colleagues whom they are fond of. There is nothing wrong with doing that, but it is important, if the feedback to be representative, also to hand out an assessment form to a colleague who is maybe not the ratee's best friend. In the theoretical framework we will return to these premises and their influence on good 360 degree feedback assessments.

Given the above, we are able to say, that when it comes to 360 degree feedback there are several premises to account for and stages that need to be attended. Summarized, it should be clear by now that 360 degree feedback assessments are highly complex material. To assess 360 degree feedback assessments thoroughly this research uses both a cross-sectional and longitudinal approach. For the longitudinal approach the different measurement points in time will be noted as 'T1' (first

measurement) and 'T2' (second measurement). With this assumption at hand the research question will be:

What explains 360 degree feedback results both on T1 and T2?

To continue this elucidation of 360 degree feedback assessment this thesis will dig deeper into the first (perception of feedback), second (acceptation of feedback) and third stage (appropriateness of feedback) presented by Ilgen et al. (1979), which are concerned with the individual's emotional and cognitive reactions that might stand in the way of performance improvement. According to the research of Brett & Atwater (2001) it appeared that individuals tend to show negative behaviour to negative feedback, but refrain from showing positive behaviour to positive feedback (Brett et al., 2001). Also here the role of the coach is emphasized. The coach should try to focus attention on the positive areas and try to minimize negative reactions. The coach should make the ratee feel positive about positive scores in order to divert attention from negative scores.

Some individuals however are able to monitor their outcomes in order to prevent negative scores. These individuals tend to monitor their own behaviour in a way to make sure others see them, the way they want to be seen. This behaviour can be described from the personality trait self-monitoring (SM). This personality trait is of major influence in creating higher scores than average according to Warech et al. (1998) and several others (Snyder, 1974; Snyder & Copeland, 1989). The definition of self-monitoring, according to Snyder (1986), is best explained as an assumption that people differ in the extent to which they *monitor* (observe, regulate and control) the public appearances of the *self* that they display in social and interpersonal relationships. In short, people influence others' opinions about who they are and behave.

Self monitoring seems to be very influential (Snyder, 1974; Snyder & Copeland, 1989), because the way one sees him or herself and the subsequent behaviour pattern can result in changing the scoring pattern. This might be caused by their self-reflecting image, which makes others very well aware of their particular behaviour (Gangestad & Snyder, 2000). The influence of self monitoring will be fully elaborated in the theoretical framework.

Based on the information above the first sub question will be:

Sub question 1: What happens with 360 degree feedback scores on T1, when an individual scores high on self monitoring personality?

In the fourth stage, the actual behavioural change and performance improvement takes place. Based on previous research there is evidence that individuals who are confronted with low scores perceive feedback most of the time as inaccurate and performance improvement does not occur (Kluger & DeNisi, 1996). In this respect, the effectiveness of 360 degree feedback assessments in the case of low feedback scores may be questioned. To find out whether performance improvement or change has occurred a follow-up assessment should be done.

Brett et al. (2001) were one the first to do an extensive follow-up assessments. Based on their results, they found evidence that when feedback was perceived as useful according to the first measurement this resulted in fewer low scores in the second measurement. Looking into the fourth stage the research question here will be:

Sub question 2: What is the relation between the scores on T1 and T2?

Some years before Brett et al. (2001) Hazucha, Hezlett & Schneider (1993) already did a small scale follow-up research to investigate changes in the skill levels of 360 degree feedback participants. They followed their participants in a two year period in order to investigate how skill development, development efforts and environmental support for development are related. One of their major findings was that 360 degree feedback was found to be useful, but guidance and coaching was found as a critical factor in improving an individual's skills. Even more important was that the coach or supervisor should provide constructive suggestions, rather than general support. Based on these findings, in combination with the previous research question, the second part of question two will be:

Sub question 3: What happens between T1 and T2, if there is constructive guidance and coaching apparent in addressing the needs of the individual's development goals?

And, finally, there seem to be a number of differences in the amount of improvement depending on the rater source or feedback giver. For instance Bernadin, Hagan & Kane (1995) found no changes in improvement when ratings were received from a supervisor or costumer, but did find improvement based on subordinate and colleague ratings. This leads us to the fourth and final subquestion.

Sub question 4: What is the influence of the rater source in improvement between T1 and T2?

Unfortunately there are only a small number of studies that took a longitudinal approach in investigating changes over time, therefore an explorative approach will be used in uncovering these questions. The aim of this research, therefore, is to identify the best focus for 360 degree feedback to maximise the value of the assessment.

In the next paragraph a further elucidation of 360-degree feedback is given and the role of self-monitoring performance is discussed. The improvement in 360-degree feedback scores in general is discussed, as is the improvement in 360-degree feedback scores in relation to guidance and supportiveness. Finally rater source will be discussed.

Theoretical Framework

This paragraph describes what 360 degree feedback is and for what purposes it can be used. Also the “Keiwijzer” assessment, which is the 360 degree feedback assessment used for this study, will be clarified. And finally each of the research questions will be discussed according to previous research and this will lead to the hypothesis and assumptions.

360 degree feedback

The central theme in this thesis is 360 degree feedback. This is a type of feedback that involves not only one person giving the feedback, but a group of people. This group consist of the director (boss, manager), other superiors, colleagues, clients or subordinates but, of course, also the ratee him or herself (Brett et al., 2001). Another description is given by Ward (1997), who introduces 360 degree feedback as a systematic collection of feedback about performance data on an individual or group level, derived from a number of stakeholders. Also important to note regarding this topic is that 360 degree feedback is often also called multi source feedback (London & Smither, 1995).

The main purpose of 360 degree feedback is a developmental one, rather than an appraisal one. This is because, for 360 degree feedback to be a success, an individual should feel free to answer without being judged on the outcomes (Brouns, 2009).

With the description and purpose discussed it is worth mentioning the group of individuals under study here and the assessment tool used. The first professional use of 360 degree feedback is recorded around the 1950s at Esso Research and Engineering Company. Since those days, 360 degree feedback has practically taken over the professional feedback market, especially in fortune 500 companies (Edwards & Ewen, 1996). But it is only since 2000s that there is also an interest in 360 degree feedback assessments on a non-managerial level. Also, for this research, a 360 degree feedback assessment will be used which is designed especially for teachers in high schools. This is still a small and undiscovered group in the world of 360 degree feedback assessments. Therefore it will be interesting to find whether previous research done mainly on a managerial level will hold up on a different group of people, namely teachers.

One of the first companies to use 360 degree feedback for other use than managerial development was B&T, a small company in the Netherlands aiming at the educational sector. Based on qualitative research done by De Jong (2000) an instrument was developed called Keiwijzer. Participants

Borsten, M.C.J. (2011). *Does 360 degree feedback work?* Master Thesis: University of Tilburg

of Keiwijzer, being mainly teachers, should hand out questionnaires to their constituencies, which consists of their supervisor, colleagues and the pupils they lecture.

Constituencies and also the individual rater self will answer questions about topics concerning the interpersonal, pedagogical, didactical, professional and organizational skills of the individual. In the research of Van der Lee (2009) a full elaboration is available of the construction and underlying constructs of Keiwijzer.

The four stages of dealing with 360 degree feedback assessments.

The four stages are concerned with what the individual does with his or her results. The first stage is the (1)perception stage, the second stage is called the (2)acceptation stage, the third stage is the stage where the individual decides on (3)appropriateness for future development and in the fourth and final stage (4)behavioural change will commence. Most often the results of 360 degree feedback assessments are presented in a graphical way. In this graph a visual representation is made of the individual's test score and the scores the individual raters gave. The raters are separated per group (colleagues and supervisor have a different visualization). Based on these graphical displays an individual will most of the time explore all the components of their 360 degree feedback results together with their feedback coach.

In this first stage; (1) "*perception stage*" several emotional and cognitive reactions might appear which, depending on the strength of these feelings/ reactions, could influence the second stage of development. These could be both negative and positive reactions and this highlights an important role for the feedback coach. If the feedback coach fails to see these emotions or, is unable to cope with the ratee's reactions, the ratee could draw the wrong conclusions based on the results. Because of the large number of constituencies, each ratee receives a highly individualized feedback report. It is very important for the coach to intensively study these reports, before discussing them with the ratee in order to make sure the coach delivers the correct message and the ratee does not make his or her own conclusions based on the feedback results. An individual who receives a report for the first time might lose the bigger picture that the assessment is trying to tell and rather focus on small details. The coach should make sure the individual does not lose sight of the main concern set out in the report. In the end, it is the ratee and the ratee alone who makes conclusions for him or herself.

The conclusions a ratee might make are part of the second stage: (2)"*acceptation of feedback*" (Ilgen et al., 1979). Based on the perceptions in stage one an individual decides whether to accept the feedback or not. It is also in this part of the feedback process that the feedback coach should be able to

discern between the positive and negative reactions of the ratee in order to guide and help the individual to understand and accept their results, rather than force the ratee to accept their results.

A positive reaction might be that the outcome will result in an increased awareness and motivation in the ratee, which could result in an improvement in their performance (Brett et al., 2001). This situation might occur when an individual is confronted with the fact that his or her constituencies gave a more or less low feedback result, but these results were accepted, rather than rejected. This confrontation might lead the individual to decide to try to improve on this competency, because he or she feels capable of scoring or deserving better on a certain competency (e.g. leadership, emotional stability).

A negative reaction might be that a low score could lead to a decrease in motivation, because the individual sees the feedback as inaccurate (Brett and Atwater, 2001, Brett, Atwater and Waldman, 2005; Festinger, 1954). Compared to the previous situation, here a defence mechanism of the individual operates his or her intentions. Instead of an open minded view towards the results in a search for recognition, here the individual tends to reject the feedback and decides not to act upon the results. This rejection could lead to an end of the process and behavioural change or performance improvement will not be established. In this situation improvement or functional use of the feedback is no longer present.

In the forgoing paragraph we have the possible situation when an individual receives a low score, but a third and final possibility is, of course, when the individual receives a high score. In short, when an individual receives a high score there is not much room to improve so improvement here is very unlikely. Whether this is in fact the case will also be part of this research.

After the perception stage and acceptation stage are attended to the process continues with the third stage in which the individual decides whether the feedback is (3) "*appropriate(ness) for future development*" (Ilgen et al., 1979). The feedback report as an accepted report is for the ratee a valuable source of information. It could be quite difficult to act upon all this information. Therefore the individual must choose what part of the information is appropriate for future development and also what part needs the most attention.

Finally, in the last stage, the individual will most of the time write a personal development plan concerned with those topics they decide to act upon in stage three. Depending on these topics, the (4) "*actual behavioural change*" is determined which, in turn could lead to performance improvement (Ilgen et al., 1979). Performance improvement is defined as the difference between the score on the first measurement (T1) and the second measurement (T2).

Now we have explained the backbones of this research, we will continue with some more specific topics of concern for this study. We will discuss the role of the self-monitoring personality and self-other rater agreement and the causes of improvement from T1 to T2.

The role of the self-monitoring personality and self-other rater agreement

After explaining the Keiwijzer assessment and the role of the four stages in the feedback process is explained, we now turn to the first concept under study, which is the self-monitoring personality. To do this, first the definition of self-monitoring performance is given and a summary of previous research concerned with self-monitoring performance. In the second part, self-other rater agreement will be discussed according to some previous research and, finally, based on this previous research, a logical interpretation will be given between these two concepts and which will lead us to the first hypothesis.

Warech et al. (1998) introduced the variable self-monitoring personality as a variable that could cause discrepancies between self and other ratings to disappear or decrease if an individual scored high on this kind of behaviour. The core of the self-monitoring is impression management according to Warech et al. (1998). An important part of impression management is working on positive social appearances (Gangestad et al., 2000). These social appearances are the reason that self-monitoring personality (SM) is discussed as a trait, rather than a behaviour. The self-monitoring personality is more an element in somebody's personality than a way somebody behaves. Warech et al. (1998) assumed that (1) high self-monitoring is associated with job-related interpersonal effectiveness and/or (2) high self-monitoring individuals are better able to control or manipulate supervisors' evaluations. For this research, the self-monitoring personality scale is chosen (rather than something like self-efficacy), because of its social and interpersonal focus. This gives a good insight into an individual's attitude, behaviour and outcomes which constitute the primary criterion domains in an organizational setting (Binning & Barret, 1989). Besides this the strong linkage between self-monitoring and impression management makes this variable especially good for work-related settings (Tedeschi & Melburg, 1984).

Snyder(1986) also studied self-monitoring personality and suggested that people differ to a certain degree in the extent to which they observe, regulate and control the appearances of the self displayed in social settings and interpersonal relationships. The results of a study by Day, Schliecher and Unckless (1998) indicate that high self-monitors have better performance ratings, receive more promotions and are more likely to emerge as leaders, whereas low self-monitors do not.

In Snyder's self-monitoring theory he distinguishes between high-self-monitors who are very sensitive and responsive to social and interpersonal cues about situationally appropriate behaviour and

low self-monitors who are less responsive to such cues. Snyder et al. (1989) linked the behavioural traits motivation and ability to 360 degree feedback and noted that self-monitors may be particularly willing and able to tailor, fashion and image their position in a way the individual assumes to match a promotion profile. In one of his papers, Snyder (1974) identified three characteristics of the high self-monitor: (1) concern for the appropriateness of social behaviour (e.g. whether looking busy is positively evaluated), (2) sensitivity to important cues (e.g. whether the boss is paying attention to you), and (3) self-regulation (e.g. if the boss is watching, act busy).

Before we are able to draw our hypothesis around self-monitoring personality, we must first explain how it might affect self-other rating agreement. For the concept of self-other rating agreement and/ or discrepancies, we turn to the work of Atwater & Yammarino (1997). This research provides an extensive overview of the different types of self-other rating agreements and/or discrepancies. For now, it is important to know that the concept of self-other rating agreement and/or discrepancies involves the comparison of the score of the self and the score of the others, where, in the case of a discrepancy, there is a big difference between the two scores.

The design of Atwater et al. (1997) consisted of five blocks. First, there are the influences that lead to a self and other rating. These influences were biodata, individual characteristics, job-relevant experiences, cognitive processes and the context within which the individual was graded. Self-monitoring personality for this matter would be part of the individual characteristics and would therefore play a role in relation to the self and other rating.

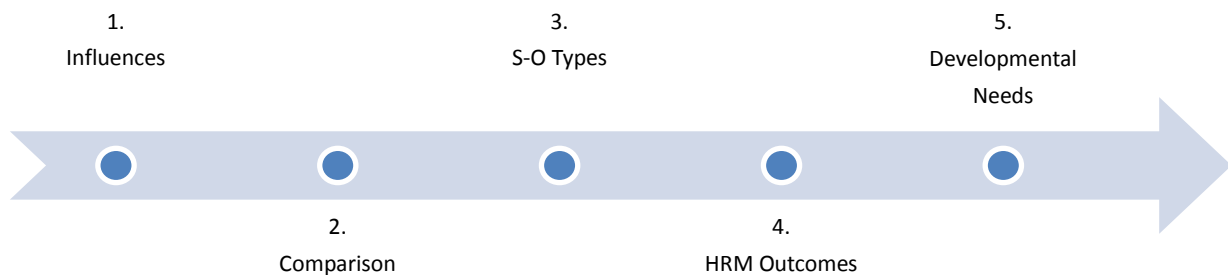


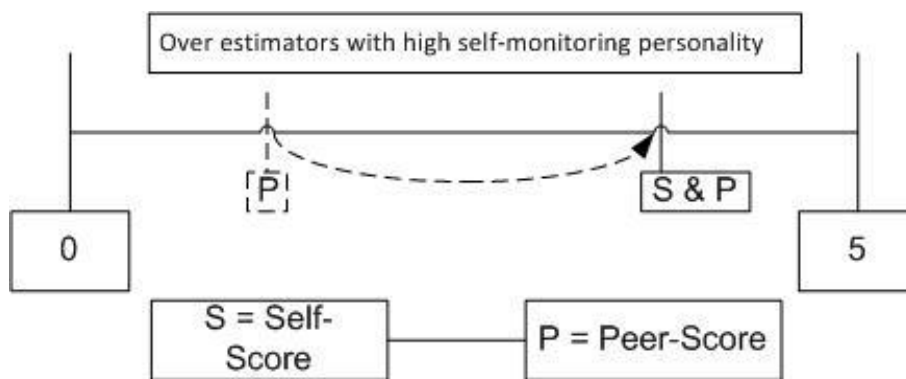
Figure 1. The self –other rating agreement process. Atwater et al. (1992)

In the second block the self and other rating are compared and based on these comparisons, Atwater et al. (1997) uncovered four potential groups (S-O Types), which are part of their third block. The first group is the in-agreement/ poor estimators group, which characterizes itself with low scores from both the self and the others. The second group is the under estimators group, which characterizes itself with high scores from others and low scores from the self. The third group is the in-agreement/

good estimators group into which individual's fall who received both high scores from themselves and from the other raters. And finally the fourth group, described as the over estimators group, is characterized with low scores from others and high scores from the self

In the fourth block Atwater et al. (1997) made a categorization based on the self-other types and the implication for the HRM outcomes (very negative, very positive, negative or mixed). Finally, in their fifth block, based on the HRM outcomes, they set the developmental needs of each group (high, low or moderate). (Appendix 2).

As discussed in the introduction based on the work of Ilgen et al. (1979) the need for improvement exists primarily for individuals who receive a low 360 score (working with scores on a 5-point basis, a low score will be defined by a score equal to or lower than two) on a certain competency. In this regard, this would imply that the research focus should be on the over-estimator group and the in-agreement/ poor estimator group, because these are the two groups that Atwater et al. (1997) defined as the groups with a high developmental need, because of the low scores on a competency. But because the in-agreement/ poor estimator group is a group where both self and other acknowledge the need for development our focus shifts towards the in-agreement/good estimators group. The in-agreement/good estimators group is the group that was defined by Atwater et al. (1997) as the group with a low development need. For the assumption made in this research, self-other rater agreement and self-monitoring personality are linked with each other. This assumption states that, because an individual shows self-monitoring behaviour, he or she is able to influence the others' scoring pattern in order to be in accordance with their own scoring.



The values indicated are examples. There can be a great variety in the exact scores

Figure 2. Over estimator with a high self-monitoring personality according to the research of Atwater et al. (1997)

This would mean that the individual without self-monitoring performance would be part of the over estimator group and would therefore have a high developmental need, whereas the individual with

a high self-monitoring personality would be part of the in-agreement/good estimator group which scores low on developmental need. This situation would result in a denial for developmental needs for these individuals, whereas the need could be just as high as those from the over estimator group, depending on the influence of the self-monitoring personality.

Therefore in this research there is a specific interest in finding out if self-monitoring personality is able to influence scores from others, because if individuals are capable of presenting themselves according to their high monitoring personality, the need for improvement would no longer be acknowledged based on test results. This brings us to our first hypothesis;

H1: Individuals who score higher on the personality trait self-monitoring will have a higher self-other rater agreement.

The causes of performance improvement from T1 to T2

Working towards the second hypothesis of this research, first we will explain performance improvement over time (T1 & T2) and the complexity of performance improvement. Then we will discuss some literature on 360 degree feedback assessment and causes for performance improvement according to the stages of Ilgen et al. (1979) and premises of Fleener et al. (2008) with a longitudinal design. Finally based on this framework we will summarize and interpret their work in this setting, which will be the ground for each of the ensuing hypothesis or assumptions.

As indicated earlier, besides the cross-sectional research on self-monitoring personality there will also be a longitudinal part to this research. For this part of the research, measurements were done on two different times, which are denominated as T1 & T2. The first point in time, when the individual has done the assessment and received his or her feedback is defined as T1. The second point in time, when an individual does his or her assessment and receives his or her feedback is defined as T2, which most of the time is at least a year later. Performance improvement will be defined as the absolute score increase between T1 and T2.

Huzacha et al. (1993) indicated that it appears to be quite difficult to determine the cause of improvement in individuals. We might be able to look at the absolute increase in score between T1 and T2 and, based on these findings conclude whether an individual improved or not, but it would be more interesting to find what caused this improvement. Most of studies indicate that the participants will improve their performance after receiving feedback, especially when the feedback is mainly presented with low scores (Smither, London & Reilly, 2005). Reilly, Smither and Vasilopoulos (1996) give two

possible reasons for this improvement. Their assumptions are built around the goal setting and control theory. Goal settings theory takes the perspective that presenting a feedback program implies to the individual that performance in the areas being measured is important (Locke and Latham, 1990). According to control theory, an individual is expected to be highly motivated to change their behaviour if their score is low in comparison to the average (Carver & Scheier, 1981). Also Smither et al. (1995) and Atwater et al. (1997) confirm that individuals receiving negative feedback are likely to exert more effort than individuals receiving positive feedback. Both theories are concerned with the individual being able to control the outcomes of their behaviour, which they like to improve.

Bearing in mind both goal setting and control theory's accounts as to why an individual might feel the drive from within to improve. The focus returns to those causes from the outside that might influence performance improvement. To begin with, we return to the work of Ilgen et al (1979) which introduced the four stages of development. What stands out in each and every stage they describe is the importance of the role of the feedback coach. The feedback coach is the individual, who together with the ratee, discusses the test results and guides the individual in the process towards writing the development plan. Therefore, the first cause of performance improvement in this research depends on the role of the feedback coach.

Also mentioned in the introduction was the research of Fleenor et al. (2008) which introduced a list of some basic premises for good 360 degree feedback assessments. One of these premises is concerned with the fact that support from the supervisor is critical for the 360 process, as well as for getting participants to set specific development goals. A second premise was that 360 degree feedback assessment should not be implemented as a stand-alone event. Both of these studies give us confirmation of the importance of the ongoing process and support after the assessments, but neither of these studied the topics from a longitudinal perspective.

It was Walker and Smither (1999) who presented a longitudinal 5-year study following feedback. Their focus was aimed at finding evidence for the use of "feedback support". They reported that managers who met with their direct supervisor to discuss their feedback improved more than other managers, and managers improved more in years when they discussed the previous year's feedback with direct supervisors than in years when they did not. Smither et al. (2004) also found that sharing multisource feedback and asking for suggestions from raters was positively related to improvement over time.

The most recent work on this specific topic is the research of Huzacha et al. (2006). In this research they touch on both premises one and two above and conclude, based on their longitudinal

research among 48 managers, that it is of great importance for an individual to receive ongoing coaching and feedback after the results of T1 were received. This was both useful for monitoring progress and preventing feedback from being a one-time event. By regularly reviewing plans and progress the feedback was a continuous process. Finally, the most important finding was that the individual who received specific and constructive guidance and coaching benefitted more from T1 to T2 compared with the individual who received more general support.

Based on the work of especially Walker et al. (1999), Smither et al. (2004) and Huzacha et al. (2006) the assumption we try to explore is that;

Assumption 1: Individuals who receive constructive guidance and coaching will experience 360 degree feedback assessment, both during and after the assessments, more positively than individuals who only received general support during the assessment phase which, in turn, might lead to more improvement.

For the second assumptions we try to explore we also take a look at the work of Fleenor et al. (2008). The third and final premise under study states that a careful selection process should be maintained in choosing and selecting raters. This premise is not directly linked to our second assumption, but emphasizes the role of the other raters.

Bernardin, Hagan & Kane (1995) tried to uncover whether the kind of group (supervisor, costumers, colleagues, subordinates) would make a difference in the level of significance the individual gives the feedback. They found evidence that no significant changes in improvement were found when ratings were received of supervisors or costumers, but did found improvement when ratings were received from subordinates and colleagues. These findings of Bernardin et al. (1995) can probably best be explained from the position of the social identity theory(SIT) and/or self-categorization theory. The SIT, which is developed by Tajfel and Turner (1986) and the Self-categorization Theory, which is developed by Turner, Hogg, Oakes, Reicher and Wetherell (1987) are related to each other in an article by Hogg and Terry (2000), which implies their relevance, also several years after their initial development. First a short description of the SIT and Self-Categorization Theory are given and after this their relation will be linked with the research of Bernardin et al. (1995) to finalize this assumption. According to Tajfel et al. (1979), SIT is described as “the individuals knowledge that he or she belongs to certain social groups together with some emotional and value significance to him or her of this group membership. Hogg et al. (2000) describe the Self-categorization theory as one that is evolved from the

SIT. They think of the self-categorization theory as that component of an extended social identity theory of the relationship between self-concept and group behavior that details the social cognitive processes that generate social identity effects. According to these two related theories we could assume that the individual that is the foci of the 360 degree feedback assessment is especially interested in the feedback of their colleagues. Therefore the second assumptions we try to defend is;

Assumption 2: Individuals who receive negative feedback from subordinates or colleagues will feel a bigger need to improve performance based on their 360 degree feedback assessments than those individuals who received low scores from supervisors or customers.

Because we have feedback results from each group of raters separately, we are also able to put the following hypothesis together.

Hypothesis 2: Individuals who receive negative feedback from colleagues will improve performance based on their 360 degree feedback assessments more than those individuals who received low scores from supervisors or customers.

In the final step of this theoretical framework we will discuss some longitudinal literature about 360 degree assessments. A problem with many of the work discussed in the introduction is the lack of a longitudinal approach therefore there is the lack of causality between the assessment on T1 and the assessment on T2 in terms of explaining improvement.

In other words, is feedback, constructive coaching and the type of group the main reason for improvement or is there something else that stimulated the occurrence of the improvement? Looking for this causality Reilly et al. (1996) did a longitudinal research in which they administered four feedback points (T1,T2,T3,T4) in a period of 2,5 years; that way they were able to ascertain whether the number of feedback moments was a cause for improvement, which could be considered as a “Hawthorne effect”, referring to the individual’s improvement being sustained merely by participating in the research. Or was there an actual improvement no matter how many times they did the research? Fortunately they were able to conclude that the improvement between T1 and T2 was actually there and they were also able to find another improvement between T3 and T4, which implies that feedback continues to improve performance even if this is shown in very small increases.

Therefore the third hypothesis will be;

H3: the lower the feedback on T1, the greater the increase in feedback on T2

Putting this paragraph together on the next page the heuristic model is displayed. Because we are not solely testing hypotheses, but also exploring assumptions it is not a conceptual model, but an heuristic model. The hypotheses 1, 2 and 3 will be tested to confirm or disconfirm, and the two assumptions will be explored, though confirming or disconfirming is not the issue. How each of these steps is taken will be explained in the next section.

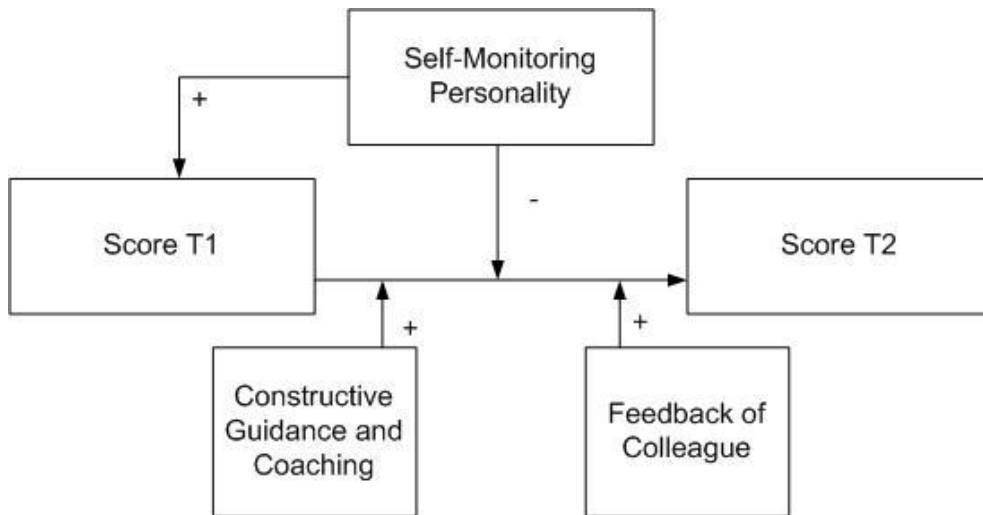


Figure 3. Heuristic Model

Method

This paragraph describes the research set-up, the sample, the procedure, the measurement scales and, finally, the analysis of this study.

Research design

This study has an exploratory character and applies a longitudinal design to measure rater scores, changes between T1 and T2, the circumstances under which performance improvement occurs and the role of self-monitoring personality in this process. For the longitudinal part of the research (H2 and H3), respondents will be used who have fulfilled the requirements of filling in Keiwijzer twice on different points in time (T1 & T2). Self-monitoring personality will be measured in a different group who has filled in Keiwijzer once before (H1). Assumption 1 and 2 were explored in a telephone interview with some of the participants of the longitudinal data set for H2.

In this study we will use Keiwijzer (2000) as the instrument to measure 360-degree feedback and collect the feedback data. This is a 360-degree feedback instrument developed by the Dutch consultancy firm Begeleiding & Training (B&T). This instrument measures the seven SBL (Stichting Beroepskwaliteit Leraren) domains of competencies that were brought together by the Dutch Foundation for Professional Teaching Competence. This foundation monitors the quality of Dutch education. Each ratee is rated by their supervisor, at least three colleagues and at least four pupils.

Procedure and Sample

It was impossible to collect enough respondents to test the entire model (fig. 2) in one sample. Therefore, an approach was followed that maximized the number of respondents per hypothesis/assumptions.

For hypothesis 1 a sample of 170 Keiwijzer participants were asked to fill out a questionnaire on SMP. 70 questionnaires were returned (41%) of which 51 (30%) could be used. Of this sample 45% is female, the average age was 45,7 years and were on average 11,51 years employed in the organization. They teach at different kind of secondary schools in the Netherlands: VMBO (29%), HAVO-VWO (25%), Scholengemeenschap (29%), Gymnasium (5%), HAVO (0%), and any other kind of secondary school (12%).

For hypothesis 2 and hypothesis 3 a sample of 62 Keiwijzer participants was drawn from the database. These participants were selected because these participants had participated in the Keiwijzer

assessment twice. Because of incomplete assessments and missing data of either T1 or T2, because the first assessments were on a non-computer basis, only 35 could be included in the final sample (56%). Of this sample 31% is female and no other information is available of these participants.

The first and second assumption are discussed based on the interviews completed with 3 individuals. These interviews have an illustrative purpose and are reported after the quantitative methods.

Measures

This research consists of several samples and therefore the following steps need to be taken for all different scales for each hypothesis. First, the different steps will be discussed and then, for each hypothesis, the scales to which these steps are applicable will be described.

For each sample, construct validity was checked by applying Principal Axis Factoring (PAF). Thereafter, the reliability of the different scales was analysed. Prior to performing Principal Axis Factoring (PAF), it was estimated whether it was allowable to conduct factor analysis on the different scales. Factor analysis was allowed as all the different scales had a Kaiser-Meyer-Okin value > 0.6 . For all, the Bartlett's Test of Sphericity was significant and the correlation matrixes showed correlations above 0.3.

For Hypothesis 1 this meant these steps needed to be followed for both the Keiwijzer assessment (Table 1 and 2; SMP) and Self-Monitoring Performance (Table 3).

For hypothesis 2 and 3 the steps above were followed for both the Keiwijzer assessment on T1 (Table 1 and 2; T1) and the Keiwijzer assessment on T2 (Table 1 and 2; T2).

360 degree feedback on T1 and T2

As mentioned in the paragraph on research design, the Keiwijzer assessment utilizes scales called SBL's (Designed by the Foundation for the Quality of Dutch Tuition). Each SBL consist of several items that measure a certain factor. The total Keiwijzer assessment consists of 150 items. Van der Lee (2009) conducted research to examine the scale validities of Keiwijzer. Based on these results, we focus on SBL 3, defined by Van der Lee (2009) as a scale that measures "Subject matter and didactic competence" and SBL5 which was defined as a scale that measures "Competent in working together in a team". The scales and the constituent items will be exactly the same for both on T1 and T2 and the analysis for Self-Monitoring Performance.

To measure “Subject matter and didactic competence” the 14 items validated by van der Lee (2009) will be used for measuring assessment results on T1, T2 and for SMP. The answer categories range from 1 = ‘Not at all’ to 5 = ‘To a high degree’. A sample question is: “Hij luistert naar nieuwe ideeën van leerlingen”.

The Scree-test of Catell and the loadings of the factor in the results of the factor analyses showed for each different rater group a similar structure. The 14 items that measure on “Subject matter and didactic competence (SBL3)” represent for each group of respondents one factor. The different factor loadings, the reliabilities and items can be found in Table 1, representing SBL 3 on T1, T2 and for the SMP sample.

Table 1: items SBL 3		T1 T2 SMP		T1 T2 SMP		T1 T2 SMP
	Self		Supervisor		Pupils	
SBL 3 Items and Factor loadings	1. Hij komt naar de interne klant toe op een manier, die door de interne klant(en) als plezierig wordt ervaren.	,788 ,447 ,409	1 Hij komt naar de interne klant toe op een manier, die door de interne klant(en) als plezierig wordt ervaren.	,903 ,747 ,678	1 Hij komt naar de interne klant toe op een manier, die door de interne klant(en) als plezierig wordt ervaren.	,581 ,676 ,716
	2 Hij luistert naar nieuwe ideeën van leerlingen.	,534 ,682 ,713	2 Hij luistert naar nieuwe ideeën van leerlingen.	,941 ,670 ,695	2 Hij luistert naar nieuwe ideeën van leerlingen.	,619 ,688 ,677
	6 Hij toont geduld naar leerlingen toe.	,426 ,495 ,491	6 Hij toont geduld naar leerlingen toe.	,871 ,721 ,772	6 Hij toont geduld naar leerlingen toe.	,572 ,562 ,679
	9 Hij geeft zijn leerlingen de ruimte voor eigen inbreng.	,601 ,590 ,744	9 Hij geeft zijn leerlingen de ruimte voor eigen inbreng.	,802 ,813 ,619	9 Hij geeft zijn leerlingen de ruimte voor eigen inbreng.	,667 ,716 ,669
	11 Hij laat aan leerlingen merken hen te begrijpen.	,542 ,738 ,647	11 Hij laat aan leerlingen merken hen te begrijpen.	,855 ,892 ,722	11 Hij laat aan leerlingen merken hen te begrijpen.	,693 ,785 ,574
	14 Hij laat merken naar leerlingen te luisteren.	,730 ,737 ,633	14 Hij laat merken naar leerlingen te luisteren.	,936 ,831 ,457	14 Hij laat merken naar leerlingen te luisteren.	,619 ,784 ,445
	17 Hij laat zien, dat hij meedenkt met leerlingen.	,749 ,799 ,676	17 Hij laat zien, dat hij meedenkt met leerlingen.	,895 ,772 ,806	17 Hij laat zien, dat hij meedenkt met leerlingen.	,656 ,676 ,773
	23 Hij luistert naar de mening van leerlingen.	,721 ,837 ,713	23 Hij luistert naar de mening van leerlingen.	,952 ,805 ,462	23 Hij luistert naar de mening van leerlingen.	,738 ,790 ,669
	24 Hij geeft leerlingen de ruimte om hun verhaal of mening te vertellen.	,769 ,788 ,732	24 Hij geeft leerlingen de ruimte om hun verhaal of mening te vertellen.	,906 ,856 ,816	24 Hij geeft leerlingen de ruimte om hun verhaal of mening te vertellen.	,658 ,720 ,757
	26 Hij laat aan de leerlingen merken, dat hij weet, wat er bij leerlingen persoonlijk leeft.	,627 ,838 ,531	26 Hij laat aan de leerlingen merken, dat hij weet, wat er bij leerlingen persoonlijk leeft.	,899 ,604 ,811	26 Hij laat aan de leerlingen merken, dat hij weet, wat er bij leerlingen persoonlijk leeft.	,636 ,592 ,723
	51 Hij legt duidelijk uit aan zijn leerlingen, wat zij leren door middel van de door hem gekozen werkvorm.	,401 ,408 ,586	51 Hij legt duidelijk uit aan zijn leerlingen, wat zij leren door middel van de door hem gekozen werkvorm.	,304 ,644 ,694	51 Hij legt duidelijk uit aan zijn leerlingen, wat zij leren door middel van de door hem gekozen werkvorm.	,577 ,612 ,757
	57 Hij haalt voor leerlingen ingewikkelde leerstof uit elkaar tot begrijpelijke stukjes lesstof.	,384 ,493 ,172	57 Hij haalt voor leerlingen ingewikkelde leerstof uit elkaar tot begrijpelijke stukjes lesstof.	,760 ,806 ,279	57 Hij haalt voor leerlingen ingewikkelde leerstof uit elkaar tot begrijpelijke stukjes lesstof.	,658 ,692 ,709
	75 Hij biedt de juiste hoeveelheid lesstof aan.	,507 ,530 ,512	75 Hij biedt de juiste hoeveelheid lesstof aan.	,763 ,835 ,669	75 Hij biedt de juiste hoeveelheid lesstof aan.	,432 ,605 ,715
	82 Hij geeft bij (groeps)opdrachten duidelijke instructies/aanwijzingen aan de leerlingen.	,461 ,537 ,528	82 Hij geeft bij (groeps)opdrachten duidelijke instructies/aanwijzingen aan de leerlingen.	,719 ,760 ,563	82 Hij geeft bij (groeps)opdrachten duidelijke instructies/aanwijzingen aan de leerlingen.	,567 ,561 ,749

SBL3 Reliability	$\alpha =$,852 ,892 ,845	$\alpha =$,961 ,944 ,924	$\alpha =$,875 ,905 ,914
------------------	------------	----------------------	------------	----------------------	------------	----------------------

To measure “Competent in working together in a team (SBL5)” 4 items validated by Van der Lee (2009) will be used for measuring assessment results on T1, T2 and for the SMP sample. The answer categories range from 1 = ‘Not at all’ to 5 = ‘To a high degree’. A sample question is: “Hij neemt initiatieven om tot afspraken binnen het docententeam te komen”.

The Scree-test of Catell and the loadings of the factor in the results of the factor analyses showed for each different rater group a similar structure. The 4 items that measure “Competent in working together in a team” represent for each group of respondents one factor. The different factor loadings, the reliabilities and items can be found in Table 2, representing SBL5 on T1, T2 and for the SMP sample.

Table 2: items SBL 5		T1 T2 SMP		T1 T2 SMP		T1 T2 SMP
	Self		Supervisor		Colleagues	
SBL 5 Items and Factor loadings	90 Hij neemt initiatieven om tot afspraken binnen het docententeam te komen.	,814 ,997 ,824	90 Hij neemt initiatieven om tot afspraken binnen het docententeam te komen.	,859 ,716 ,858	90 Hij neemt initiatieven om tot afspraken binnen het docententeam te komen.	,875 ,576 ,788
	91 Hij neemt initiatieven tot het organiseren van activiteiten, die de werksfeer ten goede komen.	,897 ,977 ,788	91 Hij neemt initiatieven tot het organiseren van activiteiten, die de werksfeer ten goede komen.	,884 ,995 ,660	91 Hij neemt initiatieven tot het organiseren van activiteiten, die de werksfeer ten goede komen.	,869 ,909 ,727
	92 Hij levert op eigen initiatief een constructieve bijdrage aan diverse overlegvormen.	,883 ,949 ,835	92 Hij levert op eigen initiatief een constructieve bijdrage aan diverse overlegvormen.	,822 ,894 ,802	92 Hij levert op eigen initiatief een constructieve bijdrage aan diverse overlegvormen.	,853 ,930 ,754
	93 Hij neemt initiatieven ter verbetering van de onderlinge werkrelaties.	,951 ,955 ,705	93 Hij neemt initiatieven ter verbetering van de onderlinge werkrelaties.	,810 ,963 ,826	93 Hij neemt initiatieven ter verbetering van de onderlinge werkrelaties.	,867 ,889 ,736
SBL5 Reliability	$\alpha =$,908 ,874 ,753	$\alpha =$,861 ,915 ,718	$\alpha =$,888 ,854 ,914

Self-monitoring Personality

To measure self-monitoring personality 16 items were used from the research of Warech et al. (1998). The items are translated in Dutch by using a back-translation procedure. The answer categories for both factors range from 1 = ‘Never true for me’ to 5 = ‘Always true for me’.

Construct validity was checked by applying Principal Axis Factoring (PAF). Thereafter, the reliability of the different scales was analysed. Prior to performing Principal Axis Factoring (PAF), it was estimated whether it was allowable to conduct factor analysis on the different scales. Factor analysis was allowed as all the different scales had a Kaiser-Meyer-Okin value > 0.6. For all, the Bartlett’s Test of Sphericity was significant and the correlation matrixes showed correlations above 0.3.

The Scree-test of Catell, and the loadings of the factors in the results of the factor analysis showed for each different rater group a similar structure, but some items were dropped from the original scale.

The first factor reflects an individual's SM motivation, which consists of 7 items. A sample question is: "Ik ben zeer gemotiveerd om anderen te beïnvloeden hoe ze over me denken".

The second factor reflects an individual's SM ability, which consists of 5 items, and is represented by items such as, "In een sociale setting, ben ik in staat mijn gedrag te veranderen als ik het idee heb dat de situatie er om vraagt". The different factor loadings, the reliabilities and items can be found in Table 3. (The Dutch translation and original English scale are both included in appendices 3 and 4).

<i>Table 3: items Self Monitoring Personality</i>	Self
Self Monitoring Motivation (SMM)	<ol style="list-style-type: none"> 1. Ik ben zeer gemotiveerd om anderen te beïnvloeden hoe ze over me denken. (.680) 2. Ik heb het gevoel dat er veel goede redenen zijn om anderen te beïnvloeden hoe zij mij zien. (.595) 3. De indrukken van anderen beïnvloeden is niet belangrijk voor mij. (.620) 4. Ik zal nooit proberen om anderen zo te leiden dat ze een bepaalde indruk van mij te laten krijgen. (.595) 5. Ik probeer niet om de indrukken die mensen van mij vormen tijdens een eerste ontmoeting te beïnvloeden. (.756) 6. Ik probeer anderen hun indrukken van mij meestal te beïnvloeden. (.642) 7. Op feestjes en andere sociale aangelegenheden, zal ik geen moeite ondernemen dingen te zeggen of doen die er voor te zorgen dat ik aardig gevonden wordt. (.584)
Self Monitoring Ability (SMA)	<ol style="list-style-type: none"> 1. Zelfs als het in mijn eigen voordeel kan werken, heb ik er toch moeite mee me anders voor te doen dan ik ben. (.717) 2. Ik ben niet bepaald goed in om er voor te zorgen dat anderen mij aardig vinden. (.655) 3. In een sociale setting, ben ik in staat mijn gedrag te veranderen als ik het idee heb dat de situatie er om vraagt. (.738) 4. Ik voel me ongemakkelijk in gezelschap van anderen en sla niet zo'n goed figuur als ik zou moeten. (.791) 5. Als ik weet wat er in een situatie van me verwacht wordt ben ik goed in staat om mijn gedrag zo aan te passen dat het past bij die situatie. (.561)
SMM Reliability	$\alpha = ,751$
SMA Reliability	$\alpha = ,699$

Qualitative Methods

For this research also three interviews were also executed. The main goal of the interviews was to uncover some information about the way individuals handle their feedback and the importance of the role of the feedback coach. Several topics were addressed.

First the individual were asked to report on how they felt about the presentation of the feedback report. Were they able to understand the graphical display of the results? Did they recognize themselves in the scores they received from the other raters and what were the most important findings for each individual?

The second topic addressed, contained questions concerning the role of the feedback coach. Individuals were asked about the number of conversations they had had and the amount of support they had received from their coach and/or supervisor.

The third topic addressed whether individuals felt supported and capable of writing their personal development plan and whether they felt as if they succeeded in attaining they goals set for themselves.

Finally, these participants were asked to give feedback for the improvement of Keiwijzer.

In the results section the outcomes of these interviews will be discussed.

Results

Correlation analysis

The first analyses conducted were the correlation analyses. These analyses were done for both the Self-monitoring personality group and the longitudinal group. These analyses were conducted to gain insight into the strength and direction of the linear relationships between the variables.

First the correlation analysis of SMP will be explored. In order to test hypotheses 1 the correlations matrix of the self-monitoring performance sample was examined. Results revealed that 13 significant correlations were found of which seven are relevant for this study. Therefore, hypothesis 1 was partially confirmed.

For this analysis four control variables were added to look for any influence other than those variables under study. The results of these analyses can be found in table 4. Although sample-size was quite small ($n=51$), we applied a level of significance of .05.

A first correlation of $(-.289^*)$ was found for the ratee's self score on SBL3 (subject matter & didactic competence) and their age, which implies that younger ratees tend to give themselves higher scores than older employees. A second correlation $(-.310^*)$ was found between the ratee's Self-Monitoring Motivation and their age, which can be understood as the younger the ratee, the higher the Self-Monitoring Motivation. A third correlation $(.277^*)$ is also concerned with our control variables and shows that the score of the supervisor giving the ratee is dependent on the level of the school. In this situation we are able to say that a supervisor gives a higher score to those individuals who teach a higher level of education than those who teach on a low level. The fourth and fifth correlations are the most interesting for studying SMP and both concerned competency on subject matter and didactic skills. The fourth correlation $(-.320^*)$ tells us that someone with a low Self-Monitoring Motivation tends to give him or herself a high score. Whereas the fifth correlation $(.320^*)$ tells us that someone with a high Self-Monitoring Ability tends to give him or herself a high score and vice versa.

Finally, we were also able to find a correlation between the rater and the two different components that we had to rate on. This means that there is a positive correlation between the score for being competent on subject matter and didactic skills and the score for being competent in working in a team given by the same individual. For the self this was $(.587^{**})$ and for the supervisor this was $(.522^{**})$.

Table 4: Correlation Matrix SMP

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Sex	1,4510	,50254	1											
2. Age	45,69	11,801	-.340*	1										
3. Tenure	11,52	11,511	-.321*	.558**	1									
4. Type of School	2,55	1,540	.010	-.155	.015	1								
5. Self SBL3	3,6484	,39669	-.164	-.289*	.148	.047	1							
6. Self SBL5	3,2172	,63868	-.083	.159	.060	.243	.587**	1						
7. Supervisor SBL3	3,8296	,41175	-.022	-.025	.089	.277*	.029	.078	1					
8. Supervisor SBL5	3,7116	,49251	-.009	.094	.212	.003	-.031	.130	.522**	1				
9. Collegua SBL5	3,7881	,45014	.118	.038	.026	-.151	.025	.200	.275	.364**	1			
10. Pupil SBL3	3,5150	,37649	.209	.078	-.106	.097	.361**	.037	.087	-.286*	.163	1		
11. SMM	2,0140	,58152	-.032	-.310*	-.108	-.037	-.320*	-.109	-.088	-.170	.003	-.124	1	
12. SMA	3,4706	,50410	-.239	.171	.186	-.128	.321*	-.015	.012	.110	-.123	.142	-.185	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

To test hypotheses 2, the correlations matrix of the longitudinal sample was examined. Very weak and insignificant correlations were found for improvement of the self, based on rater-type. These results are shown in the table below as the performance improvement of the self compared with the score received by the peers on that competence. Therefore, hypotheses 2 was not supported.

For hypothesis 3 the longitudinal sample was explored. Six correlations were found of which 5 were in the expected direction. The correlations indicated that a low score on T1 led to a bigger improvement on T2, compared with a high score on T1 and the improvement on T2. Therefore, a hypothesis 3 is largely confirmed.

For this analysis only one control variable was added to look for any influence other than those variables under study, because there were no other variables available. The results of the analysis can be found in table 5. For this sample the sample-size was even smaller (n=35), but here, also, we applied a level of significance of .05. For this correlation analysis the main interest was to find correlations between the first (T1) and second (T2) measurement. All of these were found. Of the six correlations, five were negative as expected, because a negative correlation accounted for the fact that someone with a high score improved less on T2. However, one positive correlation was found. Between the score on T1 given by the self on being competent on subject matter and didactic skills there was a correlation of (-.540). This implies that, if the individual had a high score on T1, there was a smaller improvement on T2 than when the individual had a low score on T1. This was also found for the supervisor on SBL3 (-.643**), for the supervisor on SBL5 (-.763**), for the colleagues on SBL5 (-.656**) and for the pupils on SBL3 (-.553**). The only anomaly in these correlations is the one for the self on being competent in working together in a team which gives a positive correlation of (.709**). This implies that someone

with a low score on T1 also had a low improvement towards T2, whereas the individual with a high score on T1 also had a high improvement on T2.

In the correlation analysis, four significant correlations were found that showed the same pattern as with the self-monitoring personality analysis. There was a positive correlation between the score of the self score on SBL3 and SBL5, both on T1 (.438**) and the improvement on T2 (.617**). This was also found for the supervisor but only for the improvement on T2 (.483**). This implies that a low score on being competent on subject matter and didactic skills was combined with a low score on being competent in working in a team and vice versa. Finally, for being competent in working in a team (SBL5) on T1, a correlation was also found between the colleagues and supervisors (.381*). This implies that a low score of colleagues on T1 on SBL5 was combined with a low score of supervisors on T1 on SBL5.

Table 4: Correlation Matrix Longitudinal

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Sex	1,31	,471	1												
2. T1 SBL 3 Self	3,5595	,39677	.208	1											
3. T1 SBL 5 Self	2,9742	,87915	.038	.438**	1										
4. T1SBL 3 Supervisor	3,5432	,55510	.265	.007	-.246	1									
5. T1 SBL 5 Supervisor	3,1705	,78941	.127	-.083	.231	.316	1								
6. T1 SBL 5 Colleagues	3,3012	,64422	.131	-.066	.296	.215	.381*	1							
7. T1 SBL 3 pupils	3,1208	,33849	.011	.081	.082	.165	-.187	.219	1						
8. T2 – T1 SBL 3 Self	,0271	,44540	-.119	-.540**	-.365*	.119	-.192	.032	.231	1					
9. T2 – T1 SBL 5 Self	,4917	,81985	-.031	-.466**	.709**	.106	-.288	.021	-.044	.617**	1				
10. T2 – T1 SBL 3 Sup.	-,1532	,56711	-.040	.090	.289	-.643**	-.327	.032	.115	.008	-.068	1			
11. T2 – T1 SBL 5 Sup.	,4245	,90101	-.016	.122	-.087	-.101	-.763**	-.059	.277	.271	.275	.483**	1		
12. T2 – T1 SBL 5 Coll.	,3676	,52449	.047	-.160	-.235	-.301	-.354*	-.656**	-.132	.186	.245	.318	.217	1	
13. T2 – T1 SBL 3 Pup.	,6213	,42442	.226	-.168	-.215	.230	-.032	.009	-.553**	.161	.306	-.062	.100	.177	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Results Interviews

Three individuals were willing to share their Keiwijzer experiences. Two of them were female and working in the administration department of a school in Barneveld. The other participant was male and working as a teacher somewhere in Friesland. Each of them filled in the Keiwijzer only once with a spread from two years ago to six months ago. All three of them were satisfied with the presentation of the results, but also added that the guidance of the coach was important to make the correct interpretation about the results. One of them mentioned that “... the presentation of the results was very clear...”. But she also added “... I was able to draw solid conclusions about the results with the guidance of the coach and would not have been able to do this alone...”.

All three of them recognized themselves in the results as they were presented, but both females immediately added that the questions were very difficult and had the feeling that some of their raters had misinterpreted the questions and had therefore given a score that created a discrepancy with their own score. However, the male respondent answered by saying that "... the results were no surprise and confirmed myself image".

When I asked them who the most important feedback givers were, this was different for all three. One of them found the pupils the most important, because he felt as if they were his main priority. The two ladies, being from the administration department, had no pupils as feedback givers and this group was substituted with customers. One of them gave the same line of reasoning as the male teacher did and said "... for me the customers are the most important group, because after all I work for them...", whereas her colleague found her supervisor the most important, because she said "... my supervisor is, in the end, the person who tells me what to change in my behaviour...".

All three had three meeting points with their coach. The first one was when the results were presented, the second one to help write their personal development plan and the third meeting was a check-up to discuss progress in their development goals. They were all very satisfied and positive about their feedback coach. As one of them formulated "... the coach I had was a very competent man and knew exactly what he was doing..." "... Whenever I found it difficult to put a certain feeling into words, he was capable of doing this for me...". All three acknowledged that, without the role of the coach, they would not have been able to make the progress they made. One of them however acknowledged the importance of the role of the coach in undertaking steps towards their personal development plan, but found the conversations with her supervisor in order to meet her personal development goals more important.

Finally, I wanted to know how they thought that Keiwijzer could be improved. The male participant responded by saying that "...Keiwijzer is a good instrument, but it is difficult to start a dialogue with, for instance, colleagues about the results..." "...In this situation it is very important to have a good coach that pushes you towards these conversations..." "... Within our school there are several coaches and some of them have a very active, compassionate and charismatic approach in order to get you to work on your development, but there are also coaches that do not possess these skills..." "...For me it would have been hard to develop without the support of my coach...". Both females responded that the questions were too difficult. Not only too difficult for themselves, but they also felt that their raters had trouble understanding the questions asked. One of the females felt that her

customers brought no added value to her feedback report, because she did not feel that they were in any way aware of her capabilities.

Overall, the interviews illustrated that constructive guidance and coaching is an important factor of 360 degree feedback assessments. Therefore, assumption 1 is supported. Of the three individuals interviewed they all mentioned a different rater as the most important one. Therefore, the interviews give mixed support concerning assumption 2.

Conclusion

In this final chapter a conclusion will be given based on the findings of this research and each of these conclusions will be followed by a discussion. Some limitations of this research will be discussed as well. Thereafter, recommendations for both for Keiwijzer and for 360 degree feedback assessment research in general will be discussed. This chapter will finish with an overall conclusion concerning this research.

Conclusion and discussion

First, we wanted to find evidence to find out (RQ1) what happens with 360 degree feedback scores on T1, when an individual scores high on self-monitoring. To test this, 51 participants filled in the SMP questionnaire and these scores were compared with their Keiwijzer assessments in order to test the first hypothesis that self-monitoring personality could cause biases in the feedback reports, because of the behaviour they demonstrate. No evidence was found that supported this hypothesis. None of the raters was influenced by the fact that someone had a high self-monitoring personality. The only two interesting correlations that were found showed us that someone with a low self-monitoring motivation gives him or herself a high score. This implies that an individual who does not feel any motivation to express this kind of behaviour feels better about themselves than those who are motivated to express this kind of behaviour. In contrast with this, this research also found that someone that feels capable of expressing this behaviour, rather than being motivated to express the behaviour tends to give his or herself a high score. This also implies that those who do not feel capable also rank themselves low. What we can conclude from this is that, in this research, self-monitoring personality did not cause other raters to change their scoring pattern, but did cause the self to change their scoring pattern. Therefore, self-monitoring personality is not only a variable that could influence other opinions about that person, but could also be important in self reflection. This individual behaviour caused by internal characteristics is explained by Krull (2001) as dispositional attribution. This attribution is the opposite of the outside (situational) influences that stem from the environment or culture in which that individual is found. Dispositionalism is the general tendency to prefer dispositional attribution rather than situational attribution. Therefore, self-monitoring personality, as an internal characteristic/behaviour could cause individuals to change their scoring pattern based on their dispositional attribution rather than it being caused by influences from their peers and environment (situational attribution). A final cause for why no significant results were found for self-monitoring could have to do with a topic already addressed in the

suggestions for future research by Warech et al. (1998). As already mentioned several times above it is important to use 360 degree feedback assessment as a feedback tool with a developmental purpose to be successful, but exactly because of this sheer developmental angle self-monitoring personality does not show the desired significant results. The suggestions made by several authors (Bracken, 1994; Timmreck, 1995; London et al. 1995) is that the role of self-monitoring personality will be more prevalent when this assessment tool is used for administrative decisions like, promotion and pay decisions. Because of this the individual will feel a bigger incentive to show self-monitoring behaviour in order to influence their peers and subordinates which could result in the desired promotion.

The second research question sought to uncover the relation between the 360 degree feedback scores in both T1 and T2. Specifically, based on a longitudinal research design, the research asked what happens with the score of a ratee, on T2 when he or she had received a score before on T1. Data collected from 62 individuals that participated in Keiwijzer twice were used to answer hypothesis 3. The results confirmed the hypothesis that individuals improve less if they have been given a high score on a previous assessment. This was true for all the relations on both competencies with each rater group, as well as the score of the self on being competent in working in a team. What we found here is that, if someone had given him or herself a low score on T1, most of the time the increase of someone who had given him or herself a high score was bigger than the individual that rated him or herself low.

This could be explained by the fact that the individual feels very aware of his or her role in the team and does not feel competent in improving. The results on working in a team could be so confronting that the person does not dare to rank him or herself higher on the next assessment. Nevertheless, the ranking of the other raters (colleagues and supervisor) ranking on being competent in working in a team did follow the expected scoring pattern.

The third research question aimed to find some evidence for the first assumption was concerned with the coaching an individual received after the presentation of their results. This assumption was not measured on a quantitative basis and therefore the only evidence we found was the evidence from the interviews which confirmed the assumption that constructive guidance and coaching has an important influence on the ratee's chance of improvement. The interviews took place in two different schools and guidance and coaching was available, but they noted that the way their coach operated was important for their development. The coach had to be empathetic, driven and interested in the ratee's progress in order to make constructive progress. Each ratee mentioned that the need for an active coach was important in their development process.

The fourth and final research question examined with the third hypothesis and second assumption concerned the influence of the rater source in improvement between T1 and T2. Unfortunately, the correlation matrix was unable to find any significant relations. Neither was the direction strong enough to imply any relation between the variables of performance improvement and rater source. Therefore, we can conclude that there was no significant difference whether feedback came from the supervisor, the colleagues or the pupils based on the quantitative data. Also according to the qualitative data gathered with the three interviews no strong support is available for assumption 2.

This specific research question was built around the premise that stated there should be a careful selection by the ratee in choosing their rater spectrum. Based on the work of Bernardin et al. (1995) this research assumed we would find evidence for this hypothesis on behalf of the competence concerned with didactic competence and subject matter, because the questions measuring on this competence are mainly "in-lecture" structured. Therefore, one could assume that a teacher was likely to take feedback on this competence variable less seriously, because their supervisor hardly ever attends their lectures. However, no evidence has been found. Being in a school setting, rather than a competitive managerial setting, could explain why supervisor feedback was taken just as seriously as the feedback from the other raters. Managers could feel that their supervisor is not really aware of the person that reports to them once in a while (Bernardin et al., 1995), where, in a high school setting, the atmosphere and the contacts are more person orientated, than work orientated. Also during the interviews it was clear that there was not one group of raters that was the most important. The pupils, the customers and the supervisor were all mentioned.

Limitations

The research has several limitations. Firstly, the sample size was small. For both the cross sectional and longitudinal parts there was a small number of respondents. Because Keiwijzer is a relatively new assessment, there was only a small number of people who filled in the questionnaire twice. Also, because the instrument has been subject of several changes over the past years, a large number of these people could not be included, because the questionnaire filled in the first time was not completely comparable with the questionnaire on the second point in time. Moreover, in the beginning of Keiwijzer, the questionnaire was not an online questionnaire and therefore no data of these individuals was entered in the database. Therefore the final sample this research could work with consisted of only 35 individuals. The argument for the hypothesis concerned with the increase on T2 this research might have been able to draw stronger conclusions, if the number of respondents would have

been bigger. For the other hypothesis concerned with the influence of rater type, this research would probably have found significant results, but because of the small correlations these results would not be strong enough.

Another limitation could be that control variables should be incorporated in the longitudinal design, because of a lack of knowledge of this data. Because each school has its own culture, a control variable like the type of school would be good to add to the correlation matrix in order to control for spurious relations.

Furthermore, the cross sectional sample for self-monitoring performance part had a small sample of N=51. The reason for this low response is because only a few people were willing to fill in the questionnaire about self monitoring performance. The reason why people probably refused to fill in the questionnaire is probably also the reason why no significant results were found. The specific way in which the questionnaire of Warech et al. (1998) poses the questions could cause people to answer in a socially desirable way. When confronted with a question like "I try to affect other's impressions of me most of the time", it would be reasonable to assume that people answer these questions with some resistance. People might think of themselves as bad people when answering these questions. For this research a conscious decision is made to use this variable, because it lends itself better for a work related setting, but more a "detached" questionnaire that measures, for instance, self-efficacy would have been better.

A fourth and final limitation concerned with Keiwijzer in general which is extensively discussed throughout this research is the one described by Fleenor et al. (2008) as the careful selection process of the raters in combination with the way they are approached. The raters that fill in Keiwijzer for the ratee are appointed by the ratee on free will, which could cause a positive outcome in terms of feedback, because mainly friendly colleagues and liked pupils are appointed. Besides this, these raters are invited online, which caused quite a lot of non-response and, after repeated reminders, an unrepresentative feedback. To this limitation, we can also add the fact that the ratees choose their own raters, which conflicts with an anonymity situation. Although no individual scores are presented to the individual, they did receive an averaged score from that specific rater group. Therefore, it might be better if the coach of a ratee hands out the assessments completely at random and anonymously.

Recommendations for Keiwijzer and future research

The first recommendation this research would like to make has to do with the first conclusion of this research. As stated in the conclusion, this research found evidence for all expected (negative)

relations in the increase of performance from T1 to T2 except for the self on being competent in working in a team. The explanation this research has given for the (positive) relation on this competence asks for future research with a third measurement in time to find out whether a (negative) relation can also be sustained on T3. The assumption behind this suggestion is concerned with the fact that these individuals supported by higher rater scores on T2 felt confident to improve their team working skills on this future point of assessment.

The second recommendation is also concerned with more measurements in the future. Because Keiwijzer is quite a young instrument, it will probably take some time before this is possible, but it would be very informative to do a longitudinal study like Reilly et al. (1996) with four measurement points in time to establish certainty about the continuous improvement this assessment is able to produce. Hopefully, this research will show that Keiwijzer is not a hype, with the main concern for a school to adjust to governmental regulation and after these demands are met teacher Keiwijzer loses their interest.

The third recommendation this research wants to deliver is concerned with the assumption around guidance and coaching. The qualitative part of this research was small (N=3), but also in with this small number of respondents it was clear that the way ratees receive guidance during their follow-up process was very important. Therefore, an extensive qualitative research about the influence of the coach, combined with longitudinal feedback results, could be very helpful in developing better coaches and herewith guidance for the future development of the ratees.

Based on the meta-analysis and review done in the study of Smither et al. (2005) this research's contribution is that it included an individual level variable, namely self-monitoring personality to control for performance improvement. Smither et al. (2005) also addressed the need for research into the importance of coaching. This research, although small, is able to establish some evidence that this is a good contributor in performance improvement and should therefore be more extensively researched.

The final recommendation concerning self-monitoring performance is that it would be interesting to find out, whether in a similar test case, but with the 360 degree feedback assessment used as an administrative tool for promotion or pay related decisions, the self-monitoring personality variable is able to produce significant results.

Overall conclusion

To begin with, this overall conclusion I would like to quote the work of Shute (2008) who summarizes feedback with a strong metaphor in three key words;

“...feedback might be likened to “a good murder” in that effective and useful feedback depends on three things: (a) motive (the ratee needs it), (b) opportunity (the ratee receives it in time to use it), and (c) means (the ratee is able and willing to use it). “

In my opinion these are some very good “rules” to think of before implementing any form of feedback. What is the motive of this school? Do we want our teachers to improve or do we simply want to address governmental regulations? Make sure feedback is given on time and not too late, but most of all make sure the ratee is willing to use it. In order to do that, we return to the work of both Ilgen et al. (1979) and Fleenor et al. (2008). Their articles are the returning point of this research, because they both laid down some important ground rules for effective 360 degree feedback assessments. Based on their work and that from several others which this research encountered, hopefully we can finally answer the main question: whether 360 degree feedback assessments do or do not work.

Unfortunately, there is no one conclusive answer to this question. In general, we might say that the Keiwijzer assessment works, at least for the two scales this research used. After all, an improvement was found for each of these scales according to each of the rater groups. But do 360 degree feedback assessments work in general? To give an answer to this question I would like to quote Smither et al. (2005). In their meta-analysis of 24 longitudinal studies they concluded by saying:

“We therefore think that it is time for researchers and practitioners to ask, 'Under what conditions and for whom is multisource feedback likely to be beneficial?' (rather than asking 'Does multisource feedback work?')”

This could mean that the conditions under which Keiwijzer is delivered are able to make Keiwijzer beneficial and that teachers can definitely benefit from Keiwijzer, but this does not mean that 360 degree feedback assessment in general works. 360 degree feedback assessments are highly complex, cost a lot of energy, mainly for the ratee, but also the feedback coach, raters, supervisors and probably others. They cost a lot of money and have many risks of failure. All of the above addressed

questions, premises and stages have to be successfully addressed and completed in order to have successful 360 degree feedback assessments.

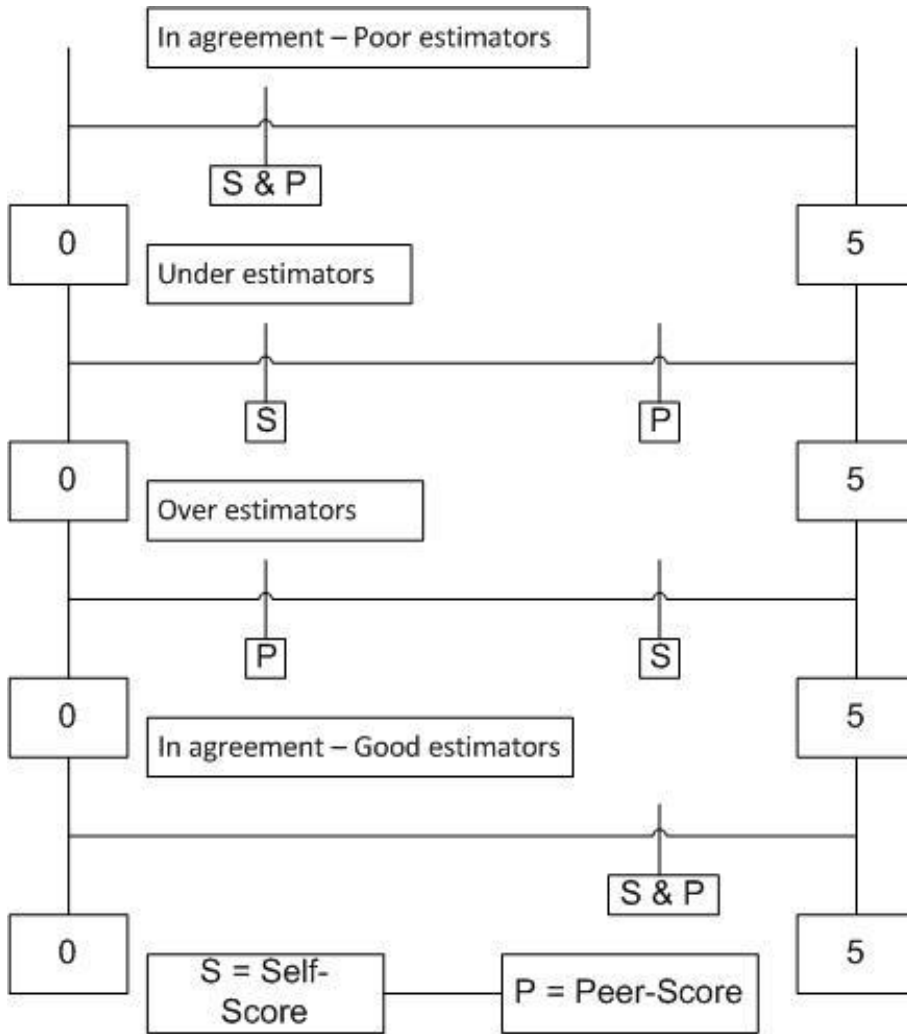
This specific research was able to establish evidence for the fact that Keiwijzer has most of their premises and stages under control, otherwise this research would probably not have found any performance improvement. We were also able to conclude that a high self-monitoring personality in this type of setting does not influence the way that raters give feedback to the ratee and therefore researchers have one fewer individual personal characteristic less to investigate as an influencer on 360 degree feedback results. Also very interesting was the fact that, although the small number of interviews was small, there was an unquestionable reaction towards the added value of the coaches.

Literature

- Atwater, L.E. & Yammarino, F.J. (1997). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, 45, 141-164.
- Bernardin, H.J., Hagan, C. & Kane, J.S. (1995). *The effects of a 360-degree appraisal system on managerial performance: No matter how cynical I get, I can't keep up*. Paper presented at the Tenth Annual Meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74, 478-494.
- Brett, J.F. & Atwater, L.E. (2001). 360-degree feedback: Accuracy, Reactions, and Perceptions of Usefulness. *Journal of Applied Psychology*, 86, 930-942.
- Brett, J.F., Atwater, L. E., & Waldman, D. A. (2005). Effective delivery of workplace discipline: Do women have to be more participatory than men. *Group and Organization Management*, 30, 487–513.
- Brouns, S.M.E. (2009). Measurement equivalence assumption of a 360 degree feedback instrument tested. Tilburg: Master Thesis
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior*. New York: Springer-Verlag.
- Day, D.V., Schleicher, D.J., & Unckless, A.I. (1996). *Self-monitoring and work-related outcomes: A meta-analysis*. Paper presented at the 11th annual Conference of the society for industrial and Organizational Psychology, San Diego, CA.
- Edwards, M.R., & Ewen, A.J. (1996). *360° Feedback: The powerful new model for Employee Assessment & performance improvement*. New York: AMACOM American Management Association.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Fleenor, J.W., Taylor, S., & Chappelow, C. (2008). Leveraging the Impact of 360-Degree Feedback. *Personnel Psychology*, 62, 175-179.
- Gandestad, S., & Snyder, M. (2000). Self-monitoring: Appraisal and reappraisal. *Psychological Bulletin*, 126, 530 – 555.
- Goodman, J. S., & Wood, R. E. (2009). Faded versus increasing feedback, task variability trajectories, and transfer of training. *Human Performance*, 22, 64-85.
- Greenhaus, J.H., Callanan, G.A., & Godshalk, V.M. (2009). *Career management*. Sage Publications Inc.
- Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management*, 32, 325-351.
- Hogg, M.A., & Terry, D.J. (2000). Social identity and self categorization processes in organizational contexts. *Academy of management review*, 25, 122-144.
- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64,349-371.
- Jong, R. De. (2000). *Competenties, kompas voor ontwikkeling*. (Doctoraalscriptie Rijksuniversiteit Groningen)
- Kluger, A.N. & DeNisi, A. (1996). The effects of feedback on performance: A historical review, a meta-analysis, and a preliminary feedback intervention. *Psychological Bulletin* 119, 254–284.

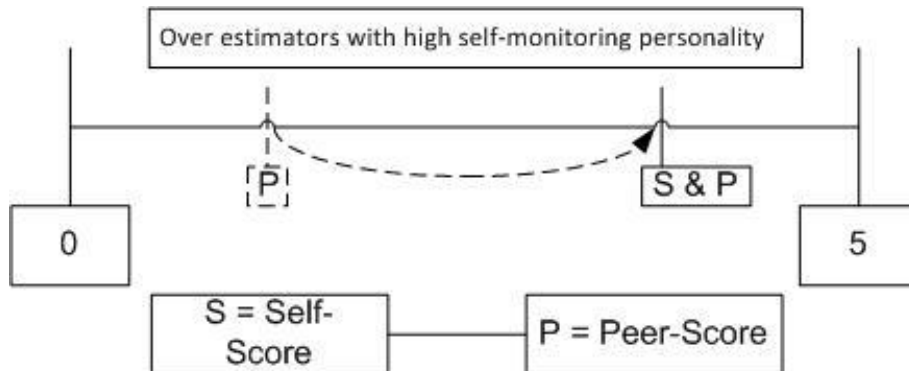
- Borsten, M.C.J. (2011). *Does 360 degree feedback work?* Master Thesis: University of Tilburg
- Krull, D.S. (2001). On partitioning the fundamental attribution error: dispositionalism and the correspondence bias. In *Cognitive Social Psychology: The Princeton Symposium on the Legacy and Future of Social Cognition*, Moskowitz GB (ed). Erlbaum: Mahwah, NJ;211–227.
- Locke, E. A. & Latham, G.P. (1990). *A Theory of Goal Setting and Task Performance*. Prentice-Hall, Englewood Cliffs, NJ.
- London, M., & Smither, J. (1995). Can multi-source feedback change perceptions of goal accomplishments, self-evaluation, and performance-related outcomes? *Personnel Psychology*, 48, 803-839.
- Luft, J. & Ingham, H. (1955). *The Johari Window: A graphic model for interpersonal relations*. University of California Western Training Laboratory.
- Reilly, R. R., Smither, J. W., & Vasilopoulos, N. L. (1996). A longitudinal study of upward feedback. *Personnel Psychology*, 49, 599-612.
- Shute, V.J. (2008). Focus on formative feedback. *Review of educational research*, 78, 153-189.
- Smither, J.W., London, M., & Reilly, R.R. (2005). Does performance improve following multisource feedback? A Theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology*, 58, 33-66.
- Smither, J.W., London, M., Vasilopoulos, N., Reilly, R.R., Millsap, R.E. & Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology*, 48, 1-34.
- Snyder, M. (1986). Public appearances, private realities: *The psychology of self-monitoring*. New York: Freeman.
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30, 526-537.
- Snyder, M. (1979). Self-monitoring processes. In L. Berkowitz (Eds.), *Advances in experimental social psychology*, 12, 85-128.
- Snyder, M. & Copeland, M. (1989). Self-monitoring processes in organizational settings. In: Giacalone, R.A. and Rosenfeld, P., Editors, 1989. *Impression management in the organization*, Erlbaum, Hillsdale, NJ, 7–19.
- Tajfel, H., & Turner, J.C. (1986). The social identity theory of intergroup behaviour. *Psychology of intergroup relations*.
- Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D. & Wetherell, M.S. (1987). Rediscovering the social group: A self-categorization theory. Basil Blackwell.
- Tedeschi, T.D., & Melburg, V. (1984). Impression Management and influence in the organization. In S.B. Bacharach & E.E. Lawler (Eds.). *Research in the sociology of organizations*, 3, 31-85.
- Lee, C. (2010). *Exploring teacher performance : the underlying structure of Keiwijzer*, Tilburg: Thesis
- Walker, A.G. & Smither, J.W. (1999). A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology*, 52, 393-423.
- Ward P. (1997) 360 Degree Feedback. *Institute of Personnel and Development, London*.
- Warech, M.A., Smither, J.W., Reilly, R.R., Millsap, R.E. & Reilly, S.P. (1998). Self-monitoring and 360-degree ratings. *Leadership Quarterly*, 9, 449–473

Appendix 1a; The possible combinations in self-other rating agreement



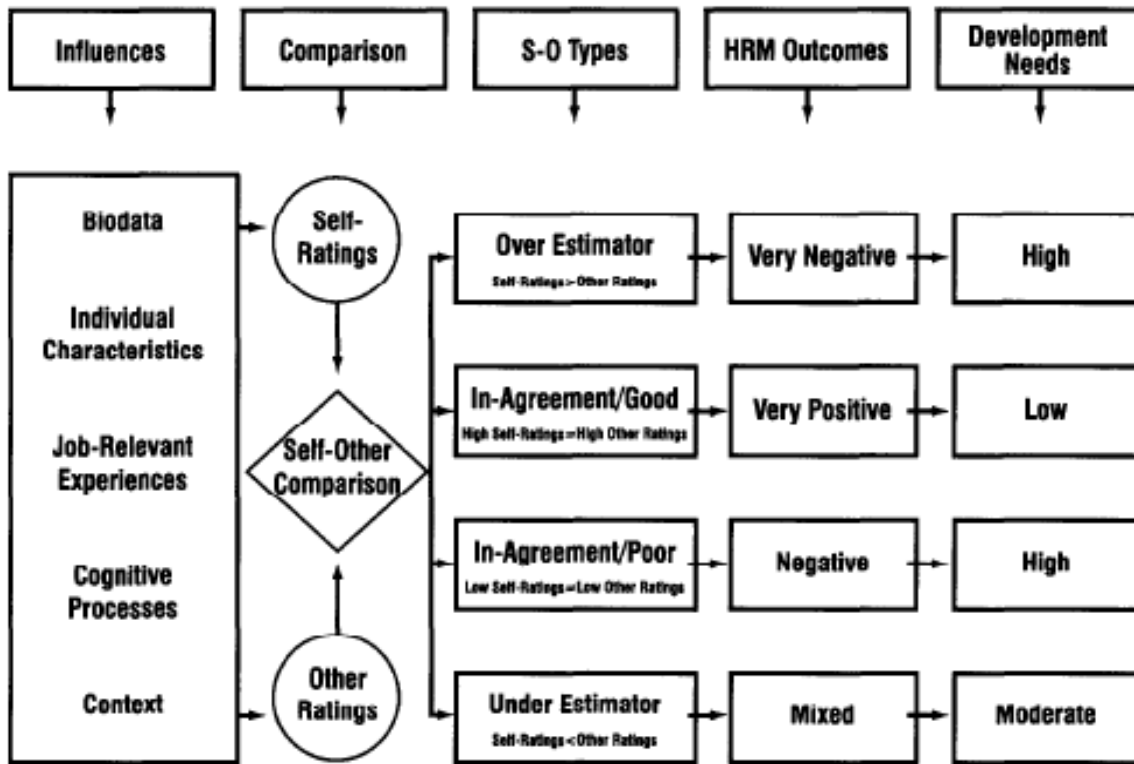
The values indicated are examples. There can be a great variety in the exact scores

Appendix 1b; The self-other rating agreement with high self-monitoring personality



The values indicated are examples. There can be a great variety in the exact scores

Appendix 2; The self-other rating agreement process



The self-other rating agreement process from: Atwater, L.E. & Yammarino, F.J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions?

Appendix 3; Self-other rater low score scales

KEIWIJZER					
Factor- Validated Multisource Items for the SBL competence framework					
Construct / source	Self *	Supervisor	Pupil	Colleague	Overlap items **
SBL 3: Subject matter and didactic competence: providing a powerful learning environment (adjusting one's teaching behaviour, one's way of reacting to the pupils, and taking individual differences into account) and encouragement of learning (directed at pupils)					
SBL 3A: <i>Adjusting behaviour to pupils</i>	23, 14, 24, 11, 2, 9, 17, 356, 1, 22,6, 20, 31, 19, 30, 16, 3, 42, 25, 41, 26, 10, 18, 107, 28 (luisteren naar leerlingen)	20, 3, 1, 11, 24, 23, 22, 14, 42, 19, 2, 26, 6, 18, 41, 9, 31, 30, 17, 356	9, 17, 24, 2, 16, 356, 23, 14, 1, 20, 11, 6, 22, 19, 26, 66	nvt	1 Hij komt naar de interne klant toe op een manier, die door de interne klant(en) als plezierig wordt ervaren. 2 Hij luistert naar nieuwe ideeën van leerlingen. 6 Hij toont geduld naar leerlingen toe. 9 Hij geeft zijn leerlingen de ruimte voor eigen inbreng. 11 Hij laat aan leerlingen merken hen te begrijpen. 14 Hij laat merken naar leerlingen te luisteren. 17 Hij laat zien, dat hij meedenkt met leerlingen. 23 Hij luistert naar de mening van leerlingen. 24 Hij geeft leerlingen de ruimte om hun verhaal of mening te vertellen. 26 Hij laat aan de leerlingen merken, dat hij weet, wat er bij leerlingen persoonlijk leeft.
SBL 3B: <i>Providing a powerful learning environment</i>	79, 74, 71, 72, 45, 80, 46, 75, 237, 76, 73, 57, 106, 82	57, 366, 58, 75, 237, 61, 76, 81, 79, 82, 80, 45, 63, 51, 103	56, 54, 50, 64, **** 368, 39, 82, 65, 51, 57, 75, 80, 69	nvt	51 Hij legt duidelijk uit aan zijn leerlingen, wat zij leren door middel van de door hem gekozen werkvorm***** 57 Hij haalt voor leerlingen ingewikkelde leerstof uit elkaar tot begrijpelijke stukjes lesstof. 75 Hij biedt de juiste hoeveelheid lesstof aan. 82 Hij geeft bij (groeps)opdrachten duidelijke instructies/aanwijzingen aan de leerlingen.
SBL 5: Competent in working together in a team: Making sure that one's work is tuned upon that of your colleagues, and contributing to the well functioning of the school organization					
	Self *	Supervisor	Pupil	Colleague	Overlap items **
	92,93,90,91, 113,105,114, 85, 112,88, 87,115, 119,109	92,90,91,93, 112,113 119,114,105.	nvt	90,91,85, 92,93,87,88 84,83	90 Hij neemt initiatieven om tot afspraken binnen het docententeam te komen. 91 Hij neemt initiatieven tot het organiseren van activiteiten, die de werksfeer ten goede komen. 92 Hij levert op eigen initiatief een constructieve bijdrage aan diverse overlegvormen. 93 Hij neemt initiatieven ter verbetering van de onderlinge werkrelaties.

*: all items have factor loadings > .3. The first item in a cell has the highest factor loading for that scale in a particular rater questionnaire.

** : Factor scale items of the different rater groups were compared on similarities and the overlapping items were selected to construct the multisource scales.

***: these items were not loading on the "pupil" scale, but because every scale was forced to exist of minimal four items, these were the next-highest loading items on the factor scales of at least two different rater groups.

****: the "pupil" factor analysis did not differentiate between SBL 1 and SBL 3B, although the self and supervisor results did indicate two distinct constructs. The items of the pupil factor were therefore divided over two different scales. This may cause higher scale correlations of these two factors for "pupils".

*****: these items were not loading on the "pupil" scale, but because every scale was forced to exist of minimal four items, these were the next-highest loading items on the factor scales of at least two different rater groups.

Appendix 4; The self-monitoring Scale

Table 1. Self-Monitoring Scale

SM Motivation (SM-M)
I am highly motivated to control how others see me. (from Leary, 1990)
I feel there are many good reasons to control how others see me. (from Leary, 1990)
Controlling others' impressions of me is not important to me. (from Leary, 1990)
In social situations, one of my goals is to get others to form a certain kind of impression of me. (from Leary, 1990)
I never try to lead others to form particular impressions of me. (from Leary, 1990)
I don't try to control the impression others form of me when I first meet them. (from Leary, 1990)
I try to affect others' impressions of me most of the time. (from Leary, 1990)
At parties and social gatherings, I do not attempt to say or do things that others will like. (from Snyder, 1974)
SM Ability (SM-A)
When I feel that the image I am portraying isn't working, I can readily change it to something that does. (from Lennox and Wolfe, 1984)
Even when it might be to my advantage, I have difficulty putting up a good front. (from Lennox and Wolfe, 1984)
I am not particularly good at making other people like me. (from Snyder, 1974)
In social situations, I have the ability to alter my behavior if I feel that something else is called for. (from Lennox and Wolfe, 1984)
I feel a bit awkward in company and do not show up quite so well as I should. (from Snyder, 1974)
I have trouble changing my behavior to suit different people and different situations. (from Snyder, 1974)
Once I know what the situation calls for, it's easy for me to regulate my actions accordingly. (from Lennox and Wolfe, 1984)
I have found that I can adjust my behavior to meet the requirements of any situation I find myself in. (from Lennox and Wolfe, 1984)

The Self-monitoring Scale from Warech, M.A., Smither, J.W., Reilly, R.R., Millsap, R.E. & Reilly, S.P. (1998). Self-monitoring and 360-degree ratings.

Appendix 5; The Dutch Self-monitoring scale

Dutch translation of the Self-monitoring Scale

Self-monitoring motivation (SM-M)

Ik ben zeer gemotiveerd om anderen te beïnvloeden in hoe ze over me denken.

Ik heb het gevoel, dat er veel goede redenen zijn om de manier waarop anderen mij zien te beïnvloeden.

De indrukken, die anderen van mij hebben beïnvloeden, is niet belangrijk voor mij.

In een sociale setting is een van mijn doelen om anderen een bepaalde indruk van mij te laten krijgen.

Ik zal nooit proberen om anderen zo te leiden, dat ze een bepaalde indruk van mij krijgen.

Ik probeer niet om de indrukken, die mensen van mij vormen, tijdens een eerste ontmoeting te beïnvloeden.

Ik probeer de indrukken van anderen over mij meestal te beïnvloeden.

Op feestjes en andere sociale aangelegenheden zal ik geen moeite ondernemen dingen te zeggen of te doen, die er voor zorgen, dat ik aardig gevonden wordt.

Self-monitoring Ability (SM-A)

Als ik het gevoel heb, dat het beeld dat ik van mezelf schets niet werkt, kan ik dat gemakkelijk veranderen naar iets dat wel werkt.

Zelfs als het in mijn eigen voordeel kan werken, heb ik er toch moeite mee me anders voor te doen dan ik ben.

Ik ben niet bepaald goed in om er voor te zorgen dat anderen mij aardig vinden.

In een sociale setting, ben ik in staat mijn gedrag te veranderen, als ik het idee heb, dat de situatie er om vraagt.

Ik voel me ongemakkelijk in gezelschap van anderen en sla niet zo'n goed figuur.

Ik heb moeite mijn gedrag aan te passen aan verschillende mensen en verschillende situaties.

Als ik weet, wat er in een situatie van me verwacht wordt, ben ik goed in staat om mijn gedrag zo aan te passen, dat het past bij die situatie.

Ik ben van mening, dat ik mijn gedrag op zo'n manier kan aanpassen, dat ik aan de gestelde eisen van de situatie, waar ik me in bevind, voldoe.

The Dutch translation of the Self-monitoring Scale from Warech, M.A., Smither, J.W., Reilly, R.R., Millsap, R.E. & Reilly, S.P. (1998). *Self-monitoring and 360-degree ratings*. Translated by Borsten, M.C.J. (2010).
Back-translation by Cartwright, A.J.(2010).