NETHERLANDS JOURNAL OF PSYCHOLOGY

VOLUME 67, NUMBER 1, JUNE 2012

Test anxiety, stress and sleep

Skill theory as a performance measure

This issue is dedicated to the memory of Wil Zeegers.

Volume 67/number 1/June 2012
Netherlands Journal of Psychology

Editorial Policy
The *Netherlands Journal of Psychology* publishes original articles of high quality, including empirical articles, review essays on selected books, theoretical and methodolological papers in any area of psychology, as well as ongoing commentaries and discussion, and short reports on the validation of psychodiagnostic instruments and related methodological tools.

The *Journal's* focus is on empirical and conceptual studies that contribute to the theoretical explanation of human behaviour and experience. Manuscripts can deal with human development, social processes and the social perspective of human behaviour, psychopathology, forensic psychology and psychiatry neuroscience, psychophysiology, philosophy of mind, the computational approach, emotion, cognition (including attention, perception, and memory), decision-making, human performance, educational psychology, health, selection and assessment and human behaviour in organisations, selection and assessment. Any other manuscripts that may be of importance to those involved with psychological research are welcomed. Manuscripts should be written to the professional and academic readership.

It is the *Netherlands Journal of Psychology's* policy to publish about recent developments in psychology and related fields. It is also the editors' policy to regularly focus on recent developments in psychology in the Netherlands. In particular, authors may be invited to contribute to special issues.

The *Netherlands Journal of Psychology* will help to focus the interest of psychologists, and others professionally interested in the field, on information about developments in a wide area of specialist research activities and results of empirical and theoretical work in the field of psychology.

The *Netherlands Journal of Psychology* offers specialists the opportunity to publish their findings to a broad audience that is scientifically interested in psychology.

The *Netherlands Journal of Psychology* aims at offering a continuous forum for intellectual discussion and transfer of knowledge and information about developments in psychology.

Categories of articles to be published and manuscripts to be welcomed are:

Review articles
Manuscripts should be limited to a maximum of 8000 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript.

Original reports of empirical research
Manuscripts should be limited to a maximum of 5000 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript.

Thematic issues
Authors with plans for planning a thematic issue are encouraged to contact the editor in an early stage. Consultation will be necessary about the number of manuscripts, number of words per manuscript, review procedure, covering of the field, etc.

Developments in the field (short)
Manuscripts should be limited to a maximum of 1000 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript.

Book review essays (short)
Manuscripts with a maximum of 1000 words need no preliminary contact with the editor. Authors planning larger contributions must contact the editor.

Letters to the editor
Manuscripts should be limited to a maximum of 500 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript.

Technical notes
Manuscripts should be limited to a maximum of 2000 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript. Authors reporting (in English) on, e.g. validation of tests and other instruments in the Dutch language or in the Netherlands are encouraged to report in *The Netherlands Journal of Psychology*. The editors want to inform non-Dutch researchers of developments and of the use of instruments in the Dutch language.

Editor's address
Professor René van Hezewijk
Faculty of Psychology, Open University of the Netherlands, PO Box 2960, 6401 DL Heerlen, the Netherlands
tel + 31 45 576 23 99
All correspondence by e-mail: rene.vanhezewijk@ou.nl

Publisher: Nederlands Instituut van Psychologen,
PO Box 2085, 3500 GB Utrecht, the Netherlands

Subscription rates
Personal rate: € 125.–
Institutional rate: € 226.–
Student rate: € 62.50
All prices are per calendar year and include value added tax (VAT)
Price per issue: € 31.95 (incl. VAT)

Subscription administration
Performis Media, PO Box 2396, 5202 CJ 's Hertogenbosch, the Netherlands, tel: + 31 73 689 58 89. For information and orders, please consult www.performis.nl

Change of address
Please notify any changes in addressee and/or address via www.performis.nl.

Payment
Please use the payment/accept giro form if possible as this simplifies the administrative process.

Advertisements
Performis Media, PO Box 2396, 5202 CJ 's Hertogenbosch, the Netherlands, tel: + 31 73 689 58 89.
The *Netherlands Journal of Psychology* is published four times a year.

ISSN 1872-552x

Contents

VOLUME 67, NUMBER 1, JUNE 2012

Relationships between adolescents' test anxiety, stress and sleep

Objective. This study aims to investigate the relationship between adolescents' test anxiety, stress and different aspects of sleep.

Method. 175 adolescents (70.8% girls, mean age 15.14 years) participated in the study. Test anxiety, stress and chronic sleep reduction were assessed at baseline using self-reports. Sleep parameters were subjectively (sleep diaries) and objectively (actigraphy) measured during the following five school nights.

Results. Stress fully mediated the relationship between test anxiety and self-reported sleep quality and test anxiety and chronic sleep reduction. Test anxiety predicted shorter subjectively and objectively measured sleep time; however, stress did not mediate these relationships.

Conclusion. We demonstrate that test anxiety is related to adolescents' stress, negatively influencing sleep quality and chronic sleep reduction, whereas the relationship between test anxiety and sleep time is not mediated by stress. Our results highlight the need to address subjective indices of sleep, especially sleep quality or chronic sleep reduction, rather than objective sleep efficiency or total sleep time when studying the effects of emotional problems on sleep.
Where: Netherlands Journal of Psychology, Volume 67, 2-8

Received 30 August 2011; Accepted 27 April 2012

Keywords: Adolescents, sleep, test anxiety, stress, chronic sleep reduction

Authors: Julia F. Dewald, Anne Marie Meijer, Frans J. Oort, and Susan M. Bögels

Research Institute of Child Development and Education, Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, the Netherlands
Correspondence to Julia F. Dewald, Research Institute of Child Development and Education, Faculty of Social and Behavioral Sciences, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, the Netherlands
E-mail: J.F.Dewald@uva.nl, tel.: +31 20 525 1280, fax: +31 20 525 1200

Test anxiety refers to the cognitive, affective and behavioural responses that accompany fear, worry and apprehension about possible failure in test situations or worry about (the consequences of) ones' poor performance (e.g. school, work) (Bögels et al., 2010). Children usually face more and more test situations as they move through the educational system, causing greater parental expectations and performance pressure, which may become internalised by the child (McDonald, 2001). With a prevalence of 20% in school children (Klingman & Papko, 1990), increasing up to 25% in high school students (Suinn, 1986), test anxiety is a prevalent anxiety in children and adolescents. Similar to other specific anxieties it was shown that the stable disposition of trait anxiety underlies test anxiety (Hodge, McCormick, & Elliott, 1997; Peleg-Popko, 2002), often causing it to become a chronic life-long condition (Bögels et al., 2010). Therefore, test anxiety is conceptualised as a situation-specific form of trait anxiety (Spielberger, 1966; Spielberger, & Vagg, 1995) and will be treated as a trait-like characteristic throughout this study.

With a prevalence of up to 45%, sleep problems (including insufficient and poor sleep) are another severe, worldwide occurring, problem in this age group (Gradisar, Gardner, & Dohnt, 2011; Liu & Zhou, 2002; Ohida et al., 2004). Based on research showing that sleep is a sensitive state that can be affected by emotions (Charuvastra & Cloitre, 2009; Mayers, Grabau, Campbell, & Baldwin, 2009), including anxiety, it can be speculated that also test anxiety negatively influences sleep.

Some evidence supporting the theoretical idea that test anxiety is related to individuals' sleep is based on research in adults demonstrating a link between (pre-sleep) rumination and worry, which are forms of cognitive arousal and important aspects of test anxiety (Cassady & Johnson, 2002), and worse self-reported sleep quality (Cropley, Dijk, & Stanley, 2006; Ohida et al., 2004; Thomsen, Mehlsen, Christensen, & Zachariae, 2003), decreased sleep efficiency (Åkerstedt, Kecklund, & Axelsson, 2007), longer sleep onset latencies (SOLs) and shorter sleep times (Kelly, 2002). Still, despite the high prevalence of adolescents' sleep problems and test anxiety and the described research in adults, no study has explicitly addressed the relationship between test anxiety and sleep in adolescents.

As test anxiety is associated with psychological and physiological arousal (Bögels et al., 2010; Hofmann, Heinrichs, & Moscovitch, 2004), it is not surprising that it often induces stress. Anxious children are characterised by a hypothalamic pituitary adrenal (HPA) axis dysregulation and significantly higher cortisol levels during the pre-sleep periods (Forbes et al., 2006), which can be interpreted as symptoms of stress. Additionally, empirical evidence from animal and human research showed that stress is closely related to objectively and subjectively measured sleep problems which can be seen in shorter sleep times and poor sleep quality (Åkerstedt, 2006; Åkerstedt et al., 2007; El-Sheikh, Buckhalt, Keller, & Granger, 2008; Fortunato & Harsh, 2006; Noland, Price, Dake, & Telljohann, 2009). In particular the anticipation of high demands or effort the next day seems to be an important component in this relationship (Åkerstedt, 2006). In general it is assumed that stress, which can be induced by test anxiety, involves increased psychological and physiological activation (e.g. HPA activity) in response to demands which seem to be incompatible with optimal sleep (Åkerstedt, 2006). Stress can therefore be seen as a state characteristic which is induced by a specific situation, event or feeling (e.g. test anxiety). Based on the above-described evidence it can be speculated that adolescents' test anxiety affects sleep through stress, meaning that stress would act as a mediator in the relationship between test anxiety and sleep.

Although subjective (e.g. sleep diaries) and objective (e.g. actigraphy) sleep assessment methods for measuring sleep duration are often correlated, subjectively experienced sleep quality and sleep efficiency, which is sometimes considered an objective measure of sleep quality, appear to be generally uncorrelated (Sadeh, 2008). These differences are not surprising as the concept of sleep quality includes subjective indices such as feeling rested when waking up, which are difficult to assess objectively. Furthermore, relations of subjectively and objectively measured sleep variables with other behavioural variables can vary considerably (Lockley, Skene, & Arendt, 1999). A recent meta-analysis showed for instance that subjective sleep quality measures were more strongly related to school performance than objective sleep measures. It has been shown that sleep variables that focus on symptoms of insufficient or poor sleep, such as sleepiness or chronic sleep reduction, may be more sensitive in detecting the impact of sleep problems on daytime functioning (Dewald, Meijer, Oort, Kerkhof, & Bögels, 2010). Therefore, this study includes objective (actigraphy) and subjective (daily sleep diaries) sleep assessment for five consecutive school nights and chronic sleep reduction as a measure of the symptoms of insufficient and/or poor sleep.

To date, research addressing the interplay between test anxiety, stress and sleep in adolescents is still lacking. Therefore, the present study aims at gaining more insight into the relationship of test anxiety, self-reported stress and different aspects of sleep (e.g. sleep quality, sleep efficiency, sleep duration, chronic sleep reduction) in adolescents. More specifically, we aim to investigate whether or not stress mediates the relationship between test anxiety and sleep.

Method

Sample

A total of 175 adolescents (70.8% girls, mean age 15.14 years) attending the third and fourth year of high school were recruited from five different schools in and around Amsterdam. Successful examination in year six offers students from these high schools the opportunity to attend university. The number of children in the family ranged from one to eight (mean: 2.29 children). The sample consisted of 66.9% families with two working parents, 25.1% families with one working parent and 1.7% families with two non-working parents (data on 5.1% were missing). The majority of the parents were married or lived together (71.4%), whereas 18.3% were divorced (5.6% of the participants reported a different family situation; data were missing for 4.7%). Most of the fathers (81.7%) and mothers (79.4%) were born in the Netherlands. More than half of our participants reported being an evening person (54.3%), whereas only 13.1% considered themselves a morning person. An overview of the descriptive statistics of test anxiety, stress and the sleep variables is presented in Table 1.

Measurements

Test anxiety: Test anxiety was measured with the test anxiety subscale of the Dutch Performance Motivation Test for Children (PMT-K (Hermans, 1983)), measuring adolescents' trait-like test anxiety. The test anxiety subscale consists of 14

Table 1 Descriptive statistics (means, SDs)

	Mean	SD
Age (years)	15.14	.56
Test anxiety	7.56	3.44
Stress	38.19	6.46
Chronic sleep reduction	35.26	6.94
Objective sleep variables (actigraphy)		
TST (hrs:min)	6:49	0:33
SOL (hrs:min)	0:22	0:17
Sleep efficiency (%)	80.94	4.79
Subjective sleep variables (sleep diary)		
Subjective sleep duration (hrs:min)	7:45	0:46
Subjective SOL (hrs:min)	0:23	0:14
Sleep quality	3.49	.61

Sleep variables refer to school nights. TST = total sleep time (actually obtained sleep); SOL = sleep onset latency (time between bedtime and sleep onset); subjective sleep duration = time between self-reported sleep onset and sleep offset

close-ended questions (e.g. ‘During a test (A) I am afraid to make mistakes, (B) I am not very afraid to make mistakes’; ‘If I know that a test is coming up I feel (A) very nervous, (B) nervous, (C) calm’; ‘During a test (A) I suffer a lot from sweating hands, (B) I suffer a little bit from sweating hands, (C) I do not suffer from sweating hands’) measuring physiological over-arousal, worry and dread about test performance. The scale is known to be a reliable and valid questionnaire with a reliability of .79 showing negative correlations with school performance (Hermans, 1983). In our study Cronbach’s alpha was .80.

Stress: Stress was measured with the Stress Questionnaire for Children (SQC) (Hartong, Krol, Maaskant, te Plate, Schuszler, & Meijer, 2003), assessing participants’ perceived stress levels. This questionnaire consists of 19 items addressing psychological and physiological stress from the previous three months (e.g. ‘I am often in a hurry’; ‘I am tense’; ‘I get easily upset’) being rated on 4-point Likert scales (1 = *this is not at all true for me*; 4 = *this is true for me*). Cronbach’s alpha in this study was .81.

Chronic sleep reduction: Chronic sleep reduction was measured with the Chronic Sleep Reduction Questionnaire (CSRQ) (Meijer, 2008). The questionnaire taps into four different symptoms of chronic sleep reduction, namely shortage of sleep (e.g. ‘I am a person who does not get enough sleep’), irritation (e.g. ‘Others think that I am easily irritated’), loss of energy (e.g. ‘Do you have enough energy during the day to do everything?’), and sleepiness (e.g. ‘Do you feel sleepy during the day?’), which are measured by 20 closed-ended questions with three ordinal response categories ranging from 1 to 3 (higher scores indicate more chronic sleep reduction). The CSRQ has been shown to be a reliable and valid measurement for chronic sleep reduction in preadolescents and adolescents (Dewald, Short, Gradisar, Oort, & Meijer, in press; Meijer, 2008). Cronbach’s alpha in this study was .86.

Subjective sleep assessment: Subjective sleep was assessed with daily sleep diaries. The following self-reported sleep variables were measured for five school nights (Sunday to Thursday night): (1) Subjective sleep duration: time between self-reported sleep onset and sleep offset; (2) Subjective sleep onset latency (subjective SOL): time between self-reported bedtime and sleep onset; and (3) Sleep quality: sleep quality was assessed by six questions about the experience of sleep quality from the previous night (e.g. ‘I felt well rested when I woke up this morning’) which were rated on 5-point Likert scales (1 = this is not at all true for me; 5 = this is true for me). These questions were adapted from a sleep quality scale, addressing aspects of falling asleep, maintaining sleep, reinitiating sleep and waking up, which has been used in previous research (Meijer & Van der Wittenboer, 2004). Aggregated scores of the sleep variables for five school nights were used for further analyses. Cronbach’s alpha in this study was .84.

Objective sleep assessment: Participants’ sleep activity was monitored using AW4 actiwatches (Cambridge Neurotechnology Ltd.). Actigraphy involves the use of a wristwatch-like portable device which can record movements over an extended period of time (e.g. a few weeks). Actigraphy is known to be a reliable and valid measure to study sleep parameters in a natural environment (Morgenthaler et al., 2007). Nocturnal activity data were logged at one minute epochs and scored with the Actiwatch Sleep Analysis 7 software. As recommended by the manufacturer we used the medium sensitivity sleep algorithm. It has been shown that this algorithm corresponds well with polysomnographic estimates (Hyde et al., 2007; Kushida, Chang, Gadkary, Guilleminault, Carrillo, & Dement, 2001). Inspection of raw actigraphy data has been recommended elsewhere (Littner et al., 2002, Sadeh & Acebo, 2002). Therefore, we visually examined all actigraphy data and corrected them if deemed necessary. Aggregated scores of the sleep variables for five school nights were used for further analyses.

Participants were instructed to wear the actigraph on their non-dominant wrist for five school nights when they went to bed and remove it in the morning after they got up. The following sleep variables were recorded: (1) Sleep onset latency (SOL): the time between somebody's attempt to fall asleep and the sleep onset time; (2) Total sleep time (TST): the number of minutes individuals actually slept, which is the time between sleep start and sleep end corrected for wake times during the sleep time; and (3) Sleep efficiency (defined as 100 x (TST/time in bed)): the percent of uninterrupted night sleep.

Procedure

The study was conducted with the approval of the University of Amsterdam Review Board. Active informed consent was obtained from participating schools, parents and participants. Demographic data, stress and trait-like test anxiety were assessed once during school times, using self-reports. After participants filled in the questionnaires, they received an actiwatch and their login information for the daily sleep diary. Sleep was monitored for five consecutive school nights (Sunday to Thursday night) with the actiwatch and subjective sleep was reported daily for the same nights in the online sleep diaries. After the actiwatches were returned participants received a summary of their measurements, were thanked and debriefed.

Statistical analyses

Data were analysed with SPSS 17.0. The dependent variables were normally distributed and all effects were controlled for the covariates gender, number of children in the family and parents' educational level and marital status. Mediation was tested following the steps by Baron and Kenney (1986). These steps consist of a series of linear regression analyses testing the relationships between the variables. First, the mediator (stress) is regressed on the independent variable (test anxiety) (criterion 1), second, the dependent variable (sleep) is regressed on the independent variable (test anxiety) (criterion 2) and third, the depending variable (sleep) is regressed on both the independent variable (test anxiety) and on the mediator (stress) (criterion 3). The preconditions of mediation are met if all relationships in the first two criteria are significant and the relationship between the mediator and the dependent variable is significant in criterion 3. Mediation is present if the relationship between the independent and dependent variable is significantly reduced after including the mediating variable and the covariate. Significance of the indirect effect was tested with the z-values of the Sobel test (Baron & Kenney, 1986).

Results

Table 2 shows the correlations between all variables. Objectively measured sleep variables (SOL, TST, sleep efficiency) and subjectively measured sleep variables (subjective SOL, subjective sleep duration, sleep quality) were correlated in the expected directions. Furthermore, we found significant correlations between subjective and objective SOL, TST and subjective sleep duration. However, sleep quality was not significantly related to any of the objective sleep variables, and sleep efficiency was significantly associated with SOL but not with any of the other subjective sleep variables. Individuals with high chronic sleep reduction scores reported longer SOLs, shorter sleep durations and worse sleep quality in their sleep diaries. However, objective SOL was the only actigraphy variable that was associated with chronic sleep reduction, meaning that correlations with TST and sleep efficiency were not significant.

The following results were found for the proposed mediation model: All required criteria (criterion 1 to 3) were met for sleep quality and chronic sleep reduction. That means that test anxiety predicted stress (criterion 1), test anxiety predicted sleep quality and chronic sleep reduction (criterion 2) and stress predicted sleep quality and chronic sleep reduction when sleep quality and chronic

Table 2 Correlations between all variables

	1	2	3	4	5	6	7	8
1. Test anxiety	-							
2. Stress	.31***	-						
3. SOL (actigraphy)	.06 n.s.	-.13 n.s	-					
4. TST (actigraphy)	-.21**	-.15 *	-.39***	-				
5. Sleep efficiency (actigraphy)	-.06 n.s.	-.30 n.s.	-.23*	.53***	-			
6. Subjective SOL (sleep diary)	-.04 n.s.	.13 n.s.	.58***	-.35**	-.57***	-		
7. Subjective sleep duration (sleep diary)	-.18*	-.15 *	-.14*	.74***	.11 n.s.	-.33***	-	
8. Sleep quality (sleep diary)	-.20**	-.47**	-.03 n.s.	.09 n.s.	.13 n.s.	-.46***	.29***	-
9. Chronic sleep reduction	.20**	.55 ***	.19*	-.13 n.s.	-.06 n.s.	.21**	-.33***	-.55***

*TST = total sleep time (actually obtained sleep); SOL = sleep onset latency (time between bedtime and sleep onset); subjective sleep duration = time between self-reported sleep onset and sleep offset. *** $p < .001$; ** $p < .01$; * $p < .05$*

Table 3 Tests of mediation

Criterion 1	β (SE)						
Test anxiety-stress	.32 (.07)***						
	Objective sleep variables (actigraphy)			**Subjective sleep variables (sleep diaries)**			
	TST β (SE)	SOL β (SE)	Sleep efficiency β (SE)	Subjective SOL β (SE)	Subjective sleep duration β (SE)	Sleep quality β (SE)	Chronic sleep reduction
Criterion 2							
Test anxiety-sleep	-.20 (.08)**	.04 (.08) n.s.	- .04 (.08) n.s.	.04 (.08) n.s.	-.15 (.07)*	-.19 (.08)*	.16 (.08)*
Criterion 3							
Mediation:							
Test anxiety-sleep/stress	-.16 (.08)*	.09 (.08) n.s.	-.04 (.08) n.s.	-.08 (.08) n.s.	-.15 (.08) n.s.	-.04 (.08) n.s.	.03 (.07) n.s.
Stress-sleep	-.14 (.08) n.s.	-.18 (.08)*	-.03 (.08) n.s.	.12 (.08) n.s.	-.10 (.08) n.s.	-.47 (.07)***	.58 (.07)***
z-value Sobel test	n.a.	n.a.	n.a.	n.a.	n.a.	-3.78***	4.00***

*TST = total sleep time (actually obtained sleep); SOL = sleep onset latency (time between bedtime and sleep onset); subjective sleep duration = time between self-reported sleep onset and sleep offset; n.a. = not applicable; *** p < .001; ** p < .01; * p < .05*

sleep reduction were regressed on both test anxiety and stress. Test anxiety was no longer significant after including stress, indicating full mediation. Furthermore, test anxiety predicted TST and subjective sleep duration; however, when TST and subjective sleep duration were regressed on both test anxiety and stress, the mediator stress did not reach significance (criterion 3). This demonstrates that test anxiety did affect TST and subjective sleep duration, but that these relationships were not mediated by stress. Table 3 shows the effects of all regression analyses.

Discussion

Stress fully mediated the relationship between adolescents' test anxiety and sleep quality and chronic sleep reduction. Additionally, test anxiety predicted TST and subjective sleep duration; however, these relationships were not mediated by stress. No significant relationships between test anxiety and subjective and objective SOL and sleep efficiency were found. Our results, showing that test anxiety predicts sleep quality, subjective sleep duration and TST, support prior studies in which pre-sleep rumination and worry were associated with worse sleep quality and shorter sleep times (Cropley et al., 2006; Ohida et al., 2004; Thomsen et al., 2003). In addition to these studies, our study gives more insight into a possible mechanism, namely stress, through which the dysfunctional relationship between test anxiety and sleep quality may be maintained. As this mediating effect was not found for TST and subjective sleep duration it can be speculated that different mechanisms play a role in sleep time and sleep quality.

Sleep quality did not correlate with any of the objective sleep variables, supporting the idea that it represents a different concept than objectively measured minutes of sleep. Furthermore, the finding that full mediation was found for sleep quality whereas effects were absent for sleep efficiency shows that individuals' subjective sleep perception may be a more accurate indicator of their daily functioning than objectively measured sleep quality (e.g. operationalised by sleep efficiency). The differences concerning the effects of sleep quality and sleep efficiency are in line with research showing that behavioural outcome variables are more strongly related to subjective than to objective sleep measurements (Dewald et al., 2010). An explanation for the stronger relationships for subjectively measured sleep variables could result from the hypothesis that the experience of test anxiety and stress may cause the misconception of poor sleep. Studies showing that individuals suffering from insomnia or depression often experience poor sleep and long SOLs, although this is often not confirmed by objective data (Voderholzer, Al- Shajlawi, Weske, Feige, & Riemann, 2003), give evidence for this assumption. However, as not only the relationship between test anxiety and sleep quality but also between test anxiety and chronic sleep reduction, (referring to symptoms of insufficient and/or poor sleep), was mediated by stress, it can be concluded that both test anxiety and stress have negative influences on individuals' sleep and daytime functioning. These results raise the idea that measures such as chronic sleep reduction may be a more sensitive assessment tool when assessing the impact of insufficient or poor sleep on individuals' daily functioning than other sleep variables such as sleep duration.

A few limitations of the study need to be addressed: First, the sample of this study represents a homogeneous group as they all attend the highest educational school level in the Netherlands. It is not unlikely that this group of students is characterised by high performance motivation, higher socio-economic status (SES), and higher IQ than students from lower educational school levels and that these characteristics may affect the variables of interest (e.g. Buckhalt, El-Sheikh, & Keller, 2007). Further research is needed in order to investigate whether or not similar relationships can be established in other samples and can therefore be generalised to the whole adolescent population. Second, other background variables that were not measured in the present study (e.g. home environment, psychopathology) may affect the relationships between test anxiety, stress and sleep. Third, our cross-sectional design does not allow conclusions about causality. Although the theoretical idea behind our study makes it likely that test anxiety causes stress and not vice versa, this should be confirmed by longitudinal data.

To summarise, we have shown that stress mediated the relationship between test anxiety and sleep quality and chronic sleep reduction. Additionally, test anxiety predicted subjective sleep duration and TST; however, these relationships were not mediated by stress. As described above, we recommend future studies to use a longitudinal design allowing causal conclusions about the relationship between test anxiety, stress and sleep. Furthermore, we used the SQC to measure stress, which is a questionnaire that refers to the previous three months. Daily stress assessment could give more insight into the actual perceived stress during the period of the study. Ecological momentary assessment (EMA) is an assessment method which involves repeated sampling of participants' current behaviours and experiences in real time in their natural environment. With this assessment method a particular event or feeling (e.g. stress) can be assessed at periodic intervals during the day. Therefore, we recommend to include such a measure in future studies in order to decrease a possible memory bias and gain even more insight into the relationship between test anxiety, stress and sleep. Finally, the study should be replicated in other samples to investigate the generalisability of the results. Based on the results of the present study demonstrating the important role of stress in the relationship between test anxiety and sleep, treatment programs that aim to reduce test anxiety should especially focus on stress when trying to improve sleep.

References

Åkerstedt, T. (2006). Psychosocial stress and impaired sleep. *Scandinavian Journal of Work and Environmental Health, 32,* 493-501.

Åkerstedt, T., Kecklund, G., & Axelsson, J. (2007). Impaired sleep after bedtime stress and worries. *Biological Psychology, 76,* 170-3.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-82.

Bögels, S. M., Alden, L., Beidel, D. C., Clark, L. A., Pine D. S., Stein, M. B., et al. (2010). Social anxiety disorder: Questions and answers for the DSM-V. *Depression and Anxiety, 27,* 168-189.

Buckhalt, J. A., El-Sheikh, M., & Keller, P. (2007). Children's sleep and cognitive functioning: race and socioeconomic status as moderators of effects. *Child Development, 78,* 213-31.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27,* 270-295.

Charuvastra, A., & Cloitre, M. (2009). Safe enough to sleep: Sleep disruptions associated with trauma, posttraumatic stress, and anxiety in children and adolescents. *Child and Adolescent Psychiatric Clinics of North America, 18,* 877-891.

Cropley, M., Dijk, D. J., & Stanley, N. (2006). Job strain, work rumination, and sleep in school teachers. *European Journal of Work and Organizational Psychology, 15,* 181-196.

Dewald, J. F., Meijer, A. M., Oort, F. J., Kerkhof, G. A., & Bögels, S. M. (2010). The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. *Sleep Medicine Reviews, 14,* 179-89.

Dewald, J. F., Short, M. A., Gradisar, M., Oort, F. J., & Meijer, A. M. (in press). The Chronic Sleep Reduction Questionnaire (CSRQ): A cross-cultural comparison and validation in Dutch and Australian adolescents. *Journal of Sleep Research.*

El-Sheikh, M., Buckhalt, J. A., Keller, P. S., & Granger, D. A. (2008). Children's objective and subjective sleep disruptions: Links with afternoon cortisol levels. *Health Psychology, 27,* 26-33.

Forbes, E. E., Williamson, D. E, Ryan, N. D., Birmaher, B., Axelson, D. A., & Dahl, R. E. (2006). Peri-sleep-onset cortisol levels in children and adolescents with affective disorders. *Biological Psychiatry, 59,* 24-30.

Fortunato, V. J., & Harsh, J. (2006). Stress and sleep quality: The moderating role of negative affectivity. *Personality and Individual Difference, 41,* 825-836.

Gradisar, M., Gardner, G., & Dohnt, H. (2011). Recent worldwide sleep patterns and problems during adolescence: a review and meta-analysis of age, region, and sleep. *Sleep Medicine, 12,* 110-8.

Hartong, I. M., Krol, M., Maaskant, A., te Plate, A., Schuszler, D., & Meijer, A. M. (2003). Psst….Slaap je al? Een onderzoek naar slaapkwaliteit. Amsterdam: University of Amsterdam.

Hermans, H. J. M. (1983). *Prestatie Motivatie test voor Kinderen PMT-K 83*. Lisse: Swets & Zeitlinger B.V.

Hodge, G. M., McCormick, J., & Elliott, R. (1997). Examination-induced distress in a public examination at the completion of secondary schooling. *British Journal of Educational Psychology, 67,* 185-197.

Hofmann, S. G., Heinrichs, N., & Moscovitch, D. A. (2004). The nature and expression of social phobia: Toward a new classification. *Clinical Psychological Review, 24,* 767-797.

Hyde, M., O'Driscoll, D. M., Binette, S., Galang, C., Tan, S. K., Verginis, N., et al. (2007). Validation of actigraphy for determining sleep and wake in children with sleep disordered breathing. *Journal of Sleep Research, 16,* 213-216.

Kelly, W. E. (2002). Worry and sleep length revisited: Worry, sleep length, and sleep disturbance ascribed to worry. *Journal of Genetic Psychology, 163*, 296-304.

Klingman, A. & Papko, A. (1990). *Test anxiety-Methods for interference and primary prevention*. In: Bar-Tal, D., & Klingman, A. (Eds.). Issues in Psychology and Counselling Education. Jerusalem: Ministry of Culture and Education, the Psychological Consulting Service.

Kushida, C. A., Chang, A., Gadkary, C., Guilleminault, C., Carrillo, O., & Dement, W. C. (2001). Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Medicine, 2,* 389-96.

Littner, M., Kushida, C. A., Anderson, W. M., Bailey, D., Berry, R. B., Davila, D. G., et al. (2003). Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. *Sleep, 26,* 337-41.

Liu, X., & Zhou, H. (2002). Sleep duration, insomnia and behavioral problems among Chinese adolescents. *Psychiatry Research, 111,* 75-85.

Lockley, S. W., Skene, D. J., & Arendt, J. (1999). Comparison between subjective and actigraphic measurement of sleep and sleep rhythms. *Journal of Sleep Research, 8,* 175-83.

Mayers, A. M., Grabau, E. A. S., Campbell, C., & Baldwin, D. S. (2009). Subjective sleep, depression and anxiety: inter-relationships in a non-clinical sample. *Human Psychopharmacology: Clinical and Experimental, 24,* 495-501.

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology, 21,* 89-101.

Meijer, A.M., & Van den Wittenboer, G. L. H. (2004). The joint contribution of sleep, intelligence and motivation to school performance. *Personality and Individual Differences, 37,* 95-106.

Meijer, A. M. (2008). Chronic sleep reduction, functioning at school and school achievement in preadolescents. *Journal of Sleep Research, 17,* 395-405.

Morgenthaler, T., Alessi, C., Friedman, L., Owens, J., Kapur, V., Boehlecke, B., et al. (2007). Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: An update for 2007. *Sleep, 30,* 519-530.

Noland, H., Price, J. H., Dake, J., & Telljohann, S. K. (2009). Adolescents' sleep behaviors and perceptions of sleep. *Journal of School Health, 79,* 224-30.

Ohida, T., Osaki, Y., Doi, Y., Tanihata, T., Minowa, M., Suzuki, K., et al. (2004). An epidemiologic study of self-reported sleep problems among Japanese adolescents. *Sleep, 27,* 978-985.

Peleg-Popko, O. (2002). Children's test anxiety and family interaction patterns. *Anxiety, Stress & Coping, 15,* 45-59.

Sadeh, A., & Acebo, C. (2002). The role of actigraphy in sleep medicine. *Sleep Medicine Reviews, 6,* 113-24.

Sadeh, A. (2008). Commentary: Comparing actigraphy and parental report as measures of children's sleep. *Journal of Pediatric Psychology, 33,* 288-302.

Spielberger, C. D. (1966). *Anxiety and behavior.* New York: Academic Press.

Spielberger, C. D., & Vagg, R. R. (1995). *Test anxiety, theory, assessment and treatment*. Washington, DC: Taylor & Francis.

Suinn, R. M. (1986). The desensitization of test-anxiety by group and individual treatment. *Behavioral Research and Therapy, 6,* 385-7.

Thomsen, D. K., Mehlsen, M. Y., Christensen, S., & Zachariae, R (2003). Rumination--relationship with negative mood and sleep quality. *Personality and Individual Differences, 34,* 1293-1301.

Voderholzer, U., Al-Shajlawi, A., Weske, G., Feige, B., & Riemann, D. (2003). Are there gender differences in objective and subjective sleep measures? A study of insomniacs and healthy controls. *Depression and Anxiety, 17,* 162-172.

JULIA F. DEWALD

PhD student at the University of Amsterdam. Her research mainly focuses on the assessment and effects of sleep problems and chronic sleep reduction in adolescents. Furthermore, she examines how adolescents' sleep problems can be improved.
E-mail: J.F.Dewald@uva.nl

DR. ANNE MARIE MEIJER

Associate professor at the University of Amsterdam. She conducts studies on childhood and parental chronic illness, and sleep (problems) of children and adolescents. Her projects focus on the relation between sleep (problems) and daytime functioning. Besides, she studies the efficacy of online treatment of adolescents with insomnia.
E-mail: A.M.Meijer@uva.nl

PROF. DR. FRANS J. OORT

Professor of Methods and Statistics in Educational Research, Director of the Graduate School of Child Development and Education, and Program Director of the Research Master Educational Sciences. His research focuses on statistical modelling, issues in measurement, and applications of statistics in psychological and educational research.
E-mail: F.J.Oort@uva.nl

PROF. DR. SUSAN M. BÖGELS

Professor in Developmental Psychopathology at the University of Amsterdam and Director of UvA Minds Academic Treatment Center for Parent and Child. Her research focuses on the transmission of anxiety from parents to children. She also studies the prevention and treatment of child anxiety disorders through child- and family-focused interventions, and mindfulness interventions for parents and children.
E-mail: S.M.Bogels@uva.nl

Application of skill theory to compare scientific reasoning of young children in different tasks

In order to study the development of scientific reasoning in children, it is necessary and also challenging to compare performances of children over different tasks and/or situations. Previous research indicates that children's performance is highly influenced by the context, which stresses the need for a task-independent measure. In this paper, skill theory (Fischer, 1980) is used as a developmental ruler. Various types of analyses of the cognitive content and structure of scientific explanations in young children are presented, which show how a coding scheme based on skill theory can provide a more thorough and multifaceted view of the child's cognitive development than standard measurements or instruments. We also demonstrate how skill theory can be used in the context-specific nature of cognitive performances and fluctuations in performance level. Skill theory promises to be a good system as a less task-specific performance measure, both in research contexts and in educational settings.

Where: Netherlands Journal of Psychology, Volume 67, 9-19

Received 10 April 2012; Accepted 30 May 2012

Keywords: Scientific reasoning, skill theory, methodology

Authors: Heidi B. Meindertsma, Marijn W. G. van Dijk, Henderien W. Steenbeek, and Paul L. C. van Geert

Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, the Netherlands
Correspondence to:
Heidi B. Meindertsma, Faculty of Behavioural and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands
E-mail address: h.meindertsma@rug.nl, tel.: + 31-50-3636451; fax: +31-50-3636304

Science education for pre-schoolers has been highly promoted in the last decade (e.g. Eshach & Fried, 2005). Studies on science education often focus on scientific reasoning. The term 'scientific reasoning' can refer to two separate types of reasoning (Crowley & Galco, 2001; Zimmerman, 2000). The first type concerns the experimentation skills of children (e.g. Klahr & Nigam, 2004; Schauble, 1996; see also the review of Zimmerman, 2000). The second form, which this article focuses on, refers to the conceptual (change in) understanding of more specific scientific phenomena. Intervention studies have focused on both kinds of reasoning. In many cases, researchers use highly specific performance measures.

In a previous study, we argued that the level of scientific reasoning and the effectiveness of an intervention (aimed at enhancing children's level of scientific reasoning) cannot be measured by studying one single task or task domain, since a task can be seen as an important contributor to the performance of the child (Meindertsma, Van Dijk, & Van Geert, submitted). In fact, we have demonstrated that different tasks result in different performance levels (for the same child, on the same day). One solution would be to use multiple tasks or transfer tasks but then the coding scheme needs to be usable for multiple tasks. However, in existing literature, a multitude of coding schemes exist that are developed for one and the same task, which complicates the comparison of different tasks and studies. As Glauert (2005) argues, there is a need for an assessment technique for both researchers and teachers. This paper therefore aims at illustrating how a performance measure based on skill theory (Fischer, 1980; Fischer & Bidell, 2006) can be used for a variety of analyses of children's scientific reasoning.

Existing performance measures

Prior studies have used a wide range of performance measures. Chien, Hsiung, and Chen

(2009) studied the effectiveness of an inquiry-based science project regarding floating and sinking and used qualitative descriptions of the explanations children gave to discuss the process of conceptual development. Glauert (2005) studied the progress of several children in a classroom context on the topic of electricity and used both a qualitative microgenetic description of the progress of the children and consequently grouped categories and types of explanations from children according to their topic.
One task for which there are a multitude of schemes is the floating and sinking task. For instance, Piaget (1930), whose pioneering work has had a great influence on the schemes developed later, stated that children go through four stages in their understanding of floating and sinking. Hsin and Wu (2011) chose to select the explanations out of all children's utterances and categorise them according to their content: weight, volume, material, water, object properties, surroundings and other. Butts, Hofman, and Anderson (1993) made a distinction between 'acceptable' and 'not yet acceptable' explanations based on the content of the explanation. The progress of the children's performance was defined as a change from giving a 'not yet acceptable' explanation in the first interview to an 'acceptable' one at the second interview. Havu-Nuutinen (2005) studied the conceptual change after an instructional process and developed a coding scheme for the given explanations about floating and sinking. They discriminated between non-relevant and non-scientific explanations, non-relational justifications and water-related justifications. The last example is derived from the study by Hardy, Jonen, Möller, and Stern (2006) which distinguishes three levels: misconceptions, explanations of everyday life and scientific explanations, which is further divided into one or more than one scientific concept.

The balance scale is another well-studied task in the literature about scientific reasoning with most studies using the rules of Siegler (1976) as a performance measure (Jansen & Van der Maas, 2002; Philips & Tolmie, 2007; Siegler & Chen, 1998). These rules are hierarchically ordered and task specific. Children using 'rule I' only focus on the weight on either side of the fulcrum while children using 'rule II' can compare the distances on both sides but only when the weight on both sides is equal. Next, children compare both weight and distance on both sides but do not yet know how to do this exactly ('rule III').'Rule IV' states that children can correctly compare both sides by multiplying weight and distance.

Few studies have compared the performance of children on different scientific tasks. Piaget (1930) qualitatively described children's reasoning for different tasks. In recent years, Tytler and Peterson (2003, 2004) used a Piagetian-based clinical interview to compare tasks such as the floating and sinking of objects, behaviour of mealworms and flight of paper whirlybirds by interviewing young children in preschool and the first year of primary school. In the 2003 study, a more quantitative analysis was done by focusing on the experimentation skills of the children. Siegler and Chen (1998) suggested that their analysis of rule learning could be used for other tasks than just the balance scale, but those tasks have to meet two requirements: 1) there must be more than one dimension and 2) children have to experience difficulty in unravelling all the separate dimensions.

The comparison of tasks is an important aspect in studying both the level of reasoning and the effect of interventions in science education, for instance in preschool education. In the Netherlands, a research program called 'Curious minds' was initiated in 2006 to promote science interest and skills in young children (Steenbeek, Van Geert, & Van Dijk, 2011; Van Benthem, Dijkgraaf, & De Lange, 2005). A range of studies have been conducted, among which those that study the effectiveness of interventions to increase teachers' teaching skills (e.g. Wetzels, Steenbeek, & Van Geert, submitted).

In order to analyse similarities and differences across tasks — in particular in the context of developmental change – these studies chose a general framework or ruler for analysing cognitive level, namely skill theory (Fischer, 1980). The importance of such a developmental ruler is further explained by Fischer and Dawson (2002). In this article we aim to show that this theory can serve as a framework for classifying explanations (which can be seen as an important part of scientific reasoning) of children, thus enabling quantitative and qualitative comparison of competence levels across tasks, situations and children (see among others Rappolt-Schlichtmann, Tenenbaum, Koepke, & Fischer, 2007). Skill theory can be used to analyse both long-term (see Van der Steen, Steenbeek & Van Geert, submitted) and short-term development. Furthermore, a process analysis can be made, which can inform us how change happens instead of describing what changes (Grannot & Parziale, 2002; Siegler, 1995; Van Geert & Steenbeek, 2005).

Dynamic skill theory

Skill theory (Fischer, 1980; Fischer & Bidell, 2006) offers a dynamic approach in describing the cognitive development of persons from infant to adult but also describes short-term learning. Skills should not be equated with internally stored algorithms, but should be viewed as patterns constructed in real-time within that specific environment. Skill theory does not state at what level a child *is* but at what level a child performs in that situation, here and now.

According to the theory, cognitive development occurs through three tiers, further divided into ten different levels. A skill at a higher level is an accumulation of lower level skills. Ages at which these levels occur have to be seen as the age at which a person, under ideal circumstances such as adequate support from a competent other person, can show this level for the first time. When confronted with a new task, a person might regress to a lower level and start building up the skill for that task (called 'scalloping', Yan & Fischer, 2002). Development can therefore be pictured as a dynamic web instead of a static ladder (Fischer & Bidell, 2006). Progress in one area does not automatically lead to progress in another area or even in another task in the same task domain. Effectiveness of an intervention on one task does not therefore inform us about the effectiveness on another task.

The context is conceived as an intrinsic part of the level of performance. In a low support context, a person will perform at a lower functional level whereas in a high support context, the optimal level of performance can be reached. The distance between the functional and optimal level is called the 'developmental range'. It is believed that it is within this range that learning takes place. When something is learned, a person first only reaches this new higher level occasionally and often falls back to the lower level, which can explain the often reported variability in performance just prior to a definite increase in level (Bassano & Van Geert, 2007; Goldin-Meadow, Alibali, & Church, 1993; Siegler & Svetina, 2002).

Complexity levels

All actions and utterances of persons can be coded according to their complexity level on an interval scale. Also, mistakes or misunderstandings get the same complexity level as their correct counterparts since they also have a complexity level and can be the starting point for more complex understandings (Schwartz & Fischer, 2004). The complexity levels are divided into three different tiers: sensorimotor, representational and abstraction (Fischer & Bidell, 2006). Each tier consists of three different levels: 'single', 'mapping' and 'system'. The highest tier, abstraction, has an additional level called 'principles'. The highest levels will not be reached until early adulthood.

The first tier is called 'sensorimotor'. Here, the child is focused on the physical sensorimotor experiences. The child does not take into account another person's point of view or understand that an object or person possesses characteristics outside the present setting. At the lowest (the single set sensorimotor) level, the child describes single sensorimotor aspects, for example that something is blue. At the next level of sensorimotor mapping, the child can also state that one object is blue and that another object is red. It combines two single sensorimotor aspects into one mapping. At the system level, consequently, the child is able to combine two mappings: the ball can bounce because it is a ball. He/she observes the bouncing of the ball and relates that to being a ball. The child understands how the mapping can be manipulated. This development roughly takes place in the first two years of life.

Fischer (1980) describes that most children first start to develop single representations in high support contexts at the age of two years. Now, the child understands that an object or person has a specific characteristic, also outside the present situation. At the first level of this tier, it is limited to one representation (e.g. 'a ball floats since it is round'). Between 3.5 and 4.5 years of age, several characteristics can be mapped together ('this ball floats since it is round and light'). The system level is generally not reached until the age of 6 or 7. At the system level two mappings can be related in one single skill which is more complex and detailed than the individual mappings ('this ball floats since it is light for its size in the water').

The abstraction tier starts with the combination of two representational systems into one system which is a single abstraction. An abstraction is conceived of as the ability to generalise ideas about an object or event outside specific situations, for example a student who understands Boyle's law (Schwartz & Fischer, 2004). According to Schwartz and Fischer, children will start to reach the abstraction level at the age of 10 to 12. The highest level, level 10, is abstract principles and is only rarely reached by highly educated or experienced people.

However, as noted before, skill levels might differ with regard to their dynamic stability. Context highly influences skill level such that a new task can result in a much lower skill level than a well-known task. Support is also important, which we have seen before in describing the 'developmental range'. Furthermore, performance at a higher level is usually accompanied by a high degree of intra-individual variability.

The present study

Until now, skill theory is used to describe performance in a variety of scientific tasks, such

as electrical circuits (Schwartz & Fischer, 2004), construction tasks (Grannot, Fischer, & Parziale, 2002) and buoyancy (Rappolt-Schlichtmann et al., 2007; Tenenbaum, Rappolt-Schlichtmann, & Zanger, 2004). Although Schwartz and Fischer (2004) described different tasks using skill theory, a comparison of performance on different tasks by the same children is hardly found in the literature.
In this paper, we aim at providing an extensive demonstration of the applicability of skill theory to analyse the performance of pre-schoolers in different scientific tasks by quantifying their explanations._ That is, we wish to demonstrate that skill theory can indeed serve as the general overarching descriptive framework for context and task-specific cognitive performance and its development.
We chose to classify the explanations of children, which provides more information than only focusing on the predictions children make (Siegler & Chen, 1998). We aim to demonstrate that by coding the explanations from children, a wide range of analyses are possible, which may answer all kinds of research questions. We will focus on short-term development in this paper, but skill theory can also be used to describe long-term development.

Method

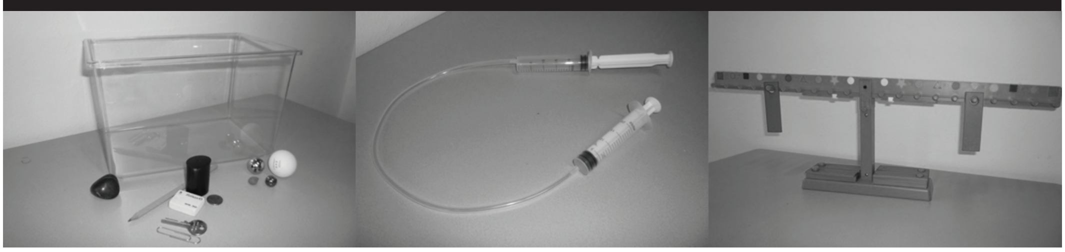
A multiple-case study was conducted, consisting of four participants (age M = 69.75 months; *SD* = 5.38), chosen from a larger database (Meindertsma, Van Dijk, Steenbeek, & Van Geert, submitted). Three boys (Lucas, David and Thomas[1]) and one girl (Kate[1]) were selected. All children attended the same urban elementary school. The ages of the children were 5;3 (Lucas), 5;8 (Kate), 6;3 (Thomas) and 6;5 (David). These particular children were chosen, since they all carried out the three tasks with one adult using the same test protocol allowing us to compare the performance of the children without being influenced by protocol or adult. The children were interviewed about three different scientific tasks (floating/sinking, linked syringes, and a balance scale) using a test protocol. The order of the tasks was randomly chosen and varied for all four children.

The aim of the interview was to elicit optimal levels of explanation without direct assistance or feedback from the adult and describe the multitude of explanations given by children. Children were presented with a new object and asked whether an object would float and sink ('will this float or sink?') and how that would be possible ('how is that possible?'). Then, the adult placed the object in the water tank ('you can give it to me, so that I can put it in the water.') and asked what had happened ('what happened?') and why this had happened ('how is that possible?'). Children were allowed to touch or manipulate the task but only within a certain, predefined limit. The adult was allowed to ask additional questions (e.g. 'what do you mean?', 'why does it float?', etcetera) if he/she thought the child could explain it better, but was not allowed to give feedback about the task or provide direct assistance.

Tasks

Water tank. Fourteen objects were placed in a water tank (Figure 1, left) during this task. The set of objects consisted of eight daily used items (match, pebble, paperclip, plastic container, coin, pencil, plastic card and eraser) and six round items (marble, clay pebble, golf ball, metal ball, ping-pong ball and apple).
Linked syringes. The linked syringes task (Figure 1, middle) is composed of two syringes connected by a plastic tube. Pushing the plunger of one of the syringes results in an outward movement of the other syringe. In the first half of the interview, the child is asked about pushing one of plungers inwards. The second half focuses on what happens when one of the plungers is pulled back.
Balance scale. The balance scale (Figure 1, right) consisted of a beam on a fixed fulcrum with on either side ten positions to hang same-weight cards on. Each position was marked by a different coloured and/or shaped figure (e.g. a red triangle, a green triangle, a blue square, etcetera). The task was designed to be increasingly difficult. There are eight, increasingly difficult, situations with the first four focusing on weight, the fifth on distance and the last three on both weight and distance.

Figure 1 The three tasks; water tank (left), linked syringes (middle) and balance scale (right)



1 Fictitious names

Table 1 Coding scheme for complexity levels based on skill theory

Code	Complexity level	Content of explanation	Water tank	Linked syringes	Balance scale
6	Representational system	Combination of all relevant representational mappings	*'The pencil is light for its size in the water'*	*'By pushing, the pressure is increased and then the air pushes the other syringe outwards'*	*'There is balance since the distance on the side with one card is twice as high as the distance on the side with two cards'*
5	Representational mapping	Two or more parts of the explaining mechanism, but the complete picture is not yet given	*'The pencil is light and thin'*	*'This thing pushes against the air and the air then pushes the other thing outwards'*	*'There is balance since there are two cards and here only one and they are on different spots'*
4	Single representation	One part of the explaining mechanism	*'The pencil is long'*	*'Because there is air'*	*'Because that card is further away'*
3	Sensorimotor system	Child states a relationship between action and result (means and end), or describes what is directly observed	*'Because it is a pencil'*	*'Because you push'*	*'Because that card is here'*

Analysis

All explanations were coded, starting at level three (sensorimotor system). When the child explained one element of the explanatory principles of the task, this was coded as level four (single representation). If the child gave two or more principles, the answers were coded as representational mapping (level five) and a complete explanation was coded as representational system (level six). Prior research (Meindertsma et al., submitted) indicated that level five was the highest level reached by pre-schoolers so the abstract tier levels were not included in this coding scheme. The coding scheme is presented in Table 1. As suggested by Schwartz and Fischer (2004), incorrect explanations were coded the same as correct explanations since they can be seen as a different pathway that can also lead to a higher level of understanding.

Reliability

Two trained coders (first and second author) independently coded the explanations of eight children (one task per child) out of the larger sample (13% of the dataset). This resulted in an adequate inter-rater reliability of .78 (kappa) with an average agreement rate of 82% (range 77-100%). Next, a Master thesis student (a third coder) was trained in using the coding scheme. This resulted in a slightly lower, but still acceptable, kappa of .71 with an average agreement rate of 74%.

Results

In this section, we will present various types of analysis of the cognitive content and structure of scientific explanations in young children and show how these can provide a more thorough and multifaceted view of the child's cognitive development than standard measurements or instruments. We will also demonstrate how skill theory can be used in the context-specific nature of cognitive performance and the fluctuations in performance level. These fluctuations should not be reduced to measurement error, but should be seen as indicative of the underlying developmental processes (Grannot et al., 2002; Van Dijk & Van Geert, 2007).

Illustration 1: Examples of explanations per task

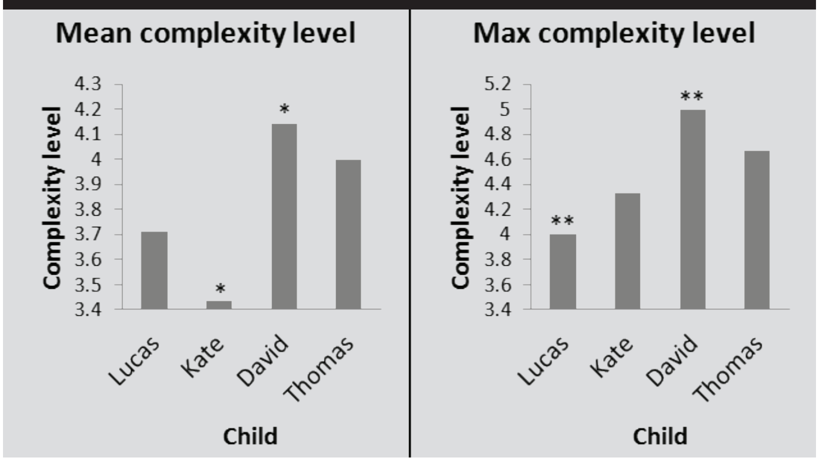
First of all, some examples of explanations and their coding according to skill theory are given. On the floating/sinking task, children using level 3 explanations referred to the name of the object (e.g. 'because it is a nut' about the clay pebble by David) or the behaviour of the object (e.g. 'because it rolled' by Kate about the marble rolling on the bottom of the water tank). One level higher, at the single representational level (level 4), children referred to specific characteristics of the object (e.g. 'because it is not so heavy' (match, Lucas), 'because it is metal' (coin, David) or 'because it is light' (plastic card, Thomas)). Not all explanations have to be consistent, e.g. Kate referred to the apple as being heavy prior to and after watching it float while she predicted it would sink. A level 5 explanation for this task might be: 'because it is lighter than the stone [pause] it could be of metal and it can also sink'. Here, David refers to both the weight and the material of the coin. Thomas claims that the metal ball is hard and heavy ('it is hard and also … heavy') and therefore sinks. For an explanation to be at level 3 on the linked syringes task, the child merely describes what happens: for instance, 'because you push it towards me' (Lucas) or 'because you pull' (Kate, David). A level 4 explanation might refer to the air (e.g. 'because the wind goes through here' (Lucas) or Kate saying: 'because it now gets air'). A level 5 explanation has to combine two level 4 explanations, so for example not only referring to air but also to the effect of the air (David: 'because air comes here and if you push it the air comes to me and that pushes this upwards').

Examples of level 3 explanations on the balance scale are 'because there is one' (Lucas points to only one side of the scale) and 'because I hang one there' (Kate). At level 4, the child only refers to weight (Kate: 'because they are equally heavy') or only to distance (David: 'they are both on the same figure'). For an explanation to be at the representational mapping level, the child must have made a combination of weight and distance, but this does not have to be precise, e.g. 'because here are three and here there are also three and they are on the same figure (David) or 'because this one has twice the value of that one' (Thomas).
In summary, by applying a skilled theoretical framework, it is possible to code a wide variety of explanations in various problem contexts in terms of a common, underlying developmental 'ruler'.

Illustration 2: Overall impression of performance

Skill theory can be used to get a general impression of the performance level of children on a specific task or over tasks. In our sample, the four children in our study gave 45 explanations: six (13%) at the level of representational mappings (level 5), 20 (44%) at the level of single representations (level 4), and ten explanations (22%) were at the level of the sensorimotor system (level 3). Nine times a child said that he or she did not know why something happened, but that was not coded as an explanation. The mean complexity level was 3.84 ($SD = .37$) and the mean maximum complexity level was 4.5 ($SD = .52$). Since the complexity level scale is at an interval level, it is possible not only to describe the explanations but also calculate means, standard deviation, etcetera. By describing these overall measures, a general impression of the performance of children on scientific tasks can be gained. Also, these measures can be used to assess the pre- or post-intervention performance level of groups of children. A more detailed look can give information about the performance per child and per task.

Figure 2 Mean complexity level (left) and maximum complexity level (right) per child



$* = p < .05$; $** = p < .1$

Illustration 3: Comparison between children

For some research questions, there is a need to know the performance level of individual children. Of the four children studied here, David (Figure 2) clearly performs at the highest level with a mean complexity level of 4.14 ($SD = .22$, $p = .03$) and a mean maximum complexity level of 5 ($SD = 0$, $p = .09$). Kate shows the lowest level of mean complexity level ($M = 3.44$, $SD = .43$, $p = .02$). For the variable maximum complexity level, Lucas has the lowest value ($M = 4.00$, $SD = .00$, $p = .09$). David and Thomas are the two six-year-olds, whereas Lucas and Kate are five years old. The same kind of analysis can be used to study age differences, but then children need to be grouped according to age.

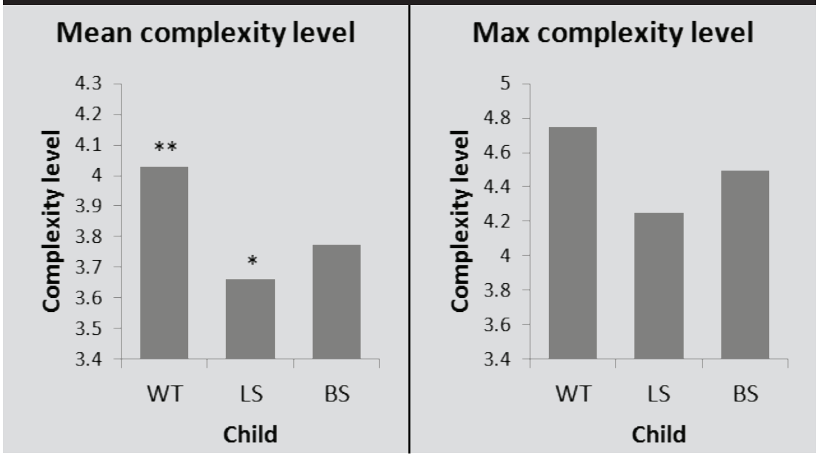
Illustration 4: Comparison of tasks

The coding scheme as presented here is especially useful for comparing the performance of (groups of) children on different tasks. In Figure 3, the performance on the three different tasks is presented. The mean complexity level of our sample group shows that these children have the most difficulty with the linked syringes task ($M = 3.66$, $SD = .32$, $p = .05$), but the floating and sinking task is responsible for the highest level of performance ($M = 4.03$, $SD = .12$, $p = .07$). The maximum level figure indicates the same distribution over the three different tasks, but this was not statistically significant. This analysis can be used to compare the difficulty of different tasks but also to analyse the effect of an intervention on a transfer task.

Illustration 5: Comparison of tasks per child

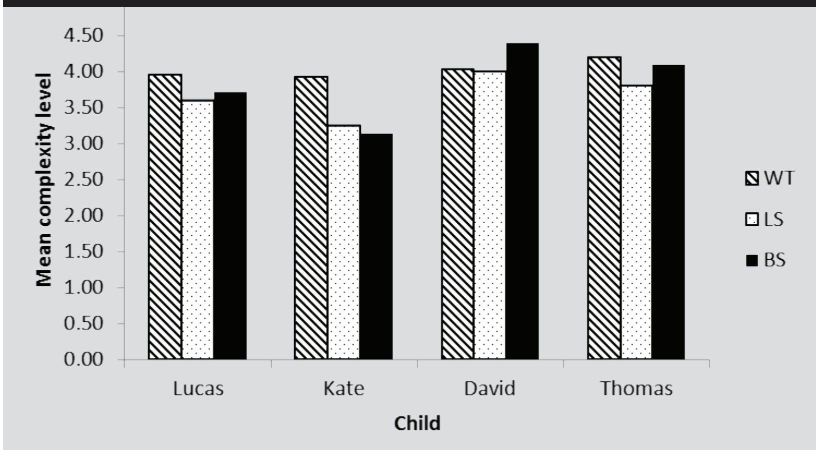
The comparison of performance on different tasks can be further specified to study the pattern per child. The performance of each child on the three different tasks is presented in Figure 4. As can be seen, Lucas ($M = 3.96$, $SD = .21$), Kate ($M = 3.93$, $SD = .37$) and Thomas ($M = 4.20$, $SD = .42$) all have their highest

Figure 3 Mean complexity level (left) and maximum complexity level (right) per task



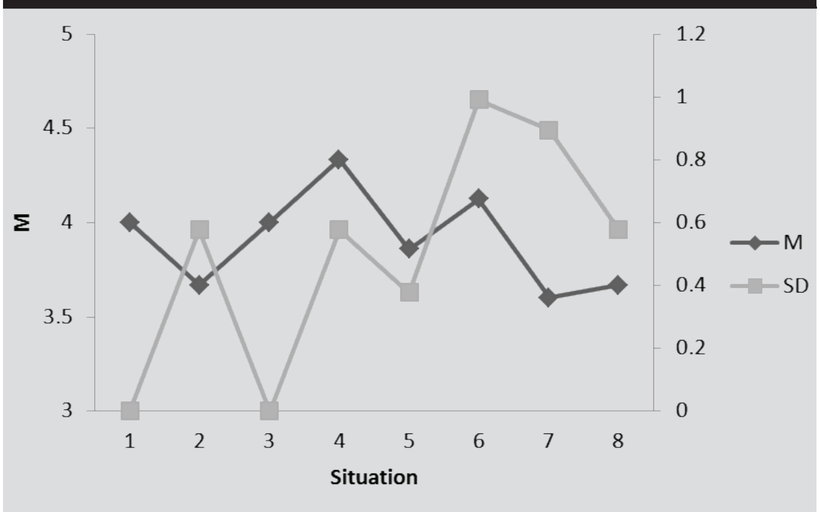
WT = water tank; LS = linked syringes; BS = balance scale;
** = p < .05, ** = p < .1*

Figure 4 Mean complexity level per child per task



WT = water tank; LS = linked syringes; BS = balance scale

Figure 5 Mean complexity level and standard deviation over time during the balance scale task



M = mean; SD = standard deviation

mean complexity level in the floating and sinking task. David, however, is best in explaining the balance scale task compared with his performance on the other two tasks ($M = 4.40$, $SD = .52$). This kind of analysis gives more detailed information about which child has more or less difficulty with one of the tasks in comparison to the performance on the other tasks. If we want to study the age effects, this kind of analysis can be used to study the development of task comprehension over different ages if we group children according to their age. For these four children, it seems that the younger children have more difficulty with the balance scale task than the older children.

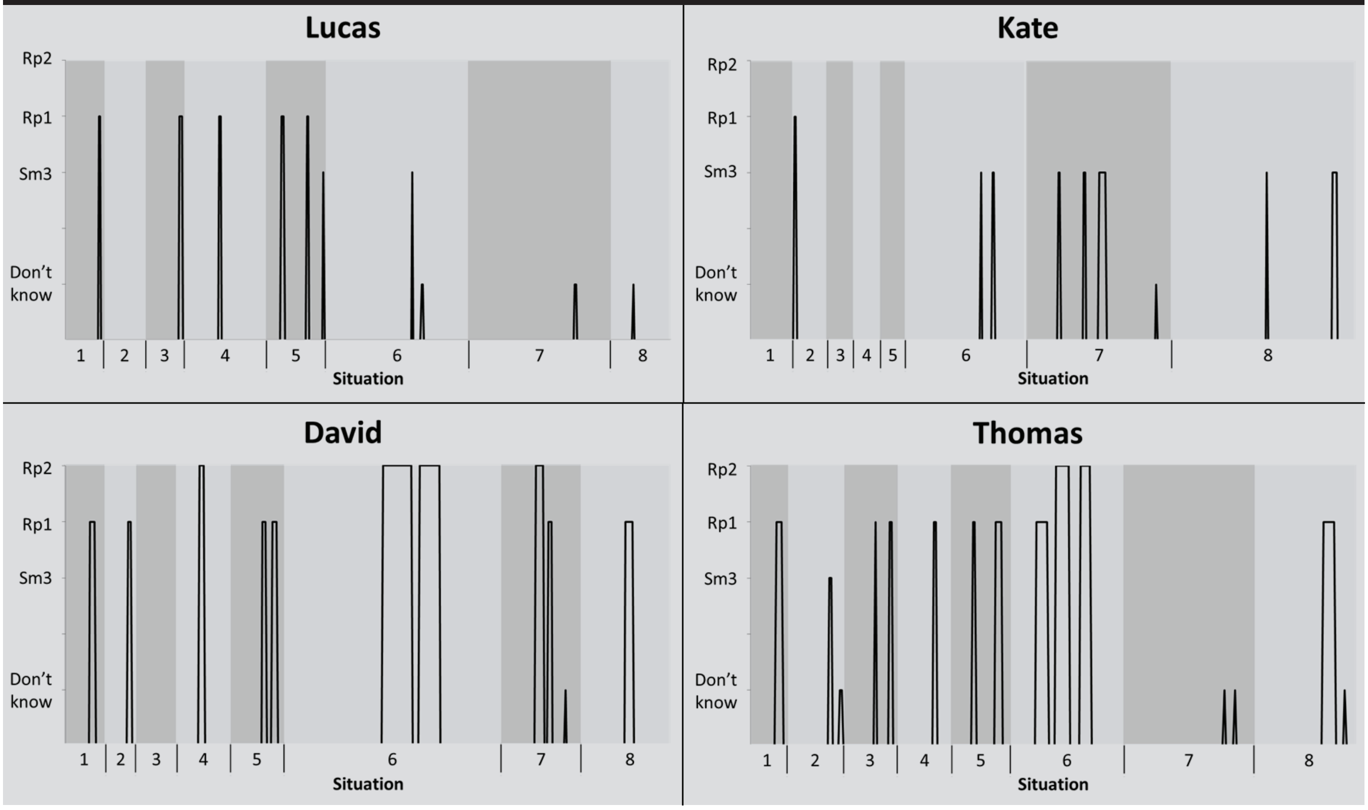
Illustration 6: Within-task performance for the whole group

Even more detailed information about the performance of a group of children can be drawn from an analysis of the performance within a task. Here, we chose the balance scale to study the short-term process within one task. In Figure 5 the mean complexity level per question is displayed for the balance scale. The mean complexity level increases from the second to the fourth question but then drops somewhat when the question about distance only is asked. The first question about weight and distance is better explained, but there is a further decline after this problem. The standard deviation increases over time, which indicates that either all children vary more or some children show a sharp decrease in complexity level. This kind of analysis can, in the case of an increasingly difficult task, give insight about at which question children drop out but also which situation children find more difficult.

Illustration 7: Within-task performance per child

Finally, the performance over time during a task can be displayed per child. In illustration 5, we compared the performance on the different tasks per child. Next, the performance during one task, the balance scale, is analysed. Figure 6 shows the level each child reaches per question during the balance scale. Since the balance scale task consists of increasingly difficult situations, a detailed analysis can give us insight into when children reach their upper limit of performance. Lucas can remain at level 4 for the first five situations which all focus on either weight or distance. In situation six, both weight and distance have to be considered which seems to be too difficult for Lucas. Kate only reaches level 4 on the second situation (about weight). David clearly can explain at the highest level. Interestingly, he already explains at level 5 on situation four which still focuses on only the weight aspect. With the last situation, he cannot maintain his level 5 explanation and falls back to level 4. For Thomas, this drop occurs at situation seven, when he claims he does not have an

Figure 6 Mean complexity level per question for the balance scale for all children



Rp = Representational; Sm = Sensorimotor

explanation. The last situation brings him back to level 4. His level 5 performance is not yet stable and a more difficult question makes him drop down to a 'don't know' answer.

By focusing on each child's own developmental trajectory, we can gain information on the knowledge level of children at each specific moment in time, which can be helpful in understanding learning processes during a task but also for educational purposes. Kate, for instance, barely explains at a level 4 for this task and has to understand that weight or distance are important aspects, whereas Lucas needs help recognising that both weight and distance matter. A teacher can focus on helping Kate understand the influence of weight or distance but offer higher level assistance to Lucas to let him realise that weight and distance both have to be taken into account. David and Thomas can be taught how weight and distance relate to each other on the balance scale.

The trajectory can also shed light on the variability or fluctuation of a child's performance. Research has indicated that variability can precede learning. In the study by Siegler (1995), a positive relation was found between variability in the pre-test and later learning. A discrepancy between verbal and nonverbal behaviour at a pre-test was also indicative for improvement during subsequent trials (Pine, Lufkin, & Messer, 2004). Variability in performance can thus indicate a window of opportunity for the teacher to help children increasing their level of understanding.

Discussion

As demonstrated in the above section, the coding scheme based on skill theory can be used for a wide range of analyses. Furthermore, related to the content, it can be reconciled with existing one-task schemes such as for the balance scale task. Rule I and II as presented by Siegler (1976) refer to children's ability to take into account weight or distance. According to skill theory, these children would be coded at level 4. One level higher (level 5 or rule III), children understand that both weight and distance are important but the exact relation between them is not yet discovered until children are using rule IV (which is equal to level six of our system). In the balance scale rules, lower levels are not described which is also comparable to our coding scheme: a level 3 answer means that children merely describe the causal relationship but do not mention an explaining mechanism. Whereas the balance scale rules only apply to tasks comparable to the balance scale (Siegler & Chen, 1998), the complexity levels of skill theory can be used in all sorts of tasks as is shown here. Our coding scheme has some additional advantages over one-task schemes and therefore promises to be a versatile and reliable system to quantify children's performance.

One of the main advantages is that it is now possible to compare the performance on different tasks due to the quantification of the performance of children. By coding the children's explanations, a child can get a mean performance level on a task, which can facilitate studying the influence of aspects such as multiple tasks or an intervention. For instance, the performance on the pre-test can be compared with the post-test. Furthermore, a transfer task can be used prior to and after the intervention to study the effect of the intervention on other tasks. Also, short-term processes can be studied to understand the *how* of the change. Long-term development can also be described by comparing children of different ages (cross-sectional research) or following children over several years (longitudinal, e.g. Van der Steen, Steenbeek, & Van Geert, submitted).

The coding scheme based on skill theory not only promises to be an efficient system as a less task-specific performance measure in research contexts. It can also help in educational settings (Parziale & Fischer, 1998) where children perform these and other scientific tasks during science classes. Teachers who are trained in using skill theory can use it as a tool to scaffold children's reasoning. Scaffolding refers to a more experienced person assisting a child to let the child reach a higher level and to the reduce this assistance when the child has reached this new level (Grannot et al., 2002). Ideally, the level of assistance is at an optimal higher level than the current child's level, which results in a maximum learning effect (Van Geert & Steenbeek, 2005). A trained teacher might ask a child playing with a task if he or she can explain what happens. The teacher can then analyse the skill level of the child instantly and ask questions or offer support so that the child can combine current skills to construct a higher skill level. By using skill theory, teachers do not only enhance the performance of the children, their own capabilities might also be expanded. By adjusting the coding scheme for a specific task, the teacher gains a better understanding of the task which in turn can increase his/her confidence in science teaching and will subsequently influence the teaching skills (Wetzels et al., 2011).

In our coding scheme, we focused on explanations from children. This means that children have to verbalise their thoughts, which might not be equally easy for all children. However, according to Siegler and Chen (1998), children are capable of reliably describing their used strategies immediately after solving a problem so we assume this should not hinder the use of skill theory. Although we only coded explanations, skill theory can also be used to code every task-related utterance of a person (Van der Steen et al., submitted) as well as nonverbal actions. Prior to the ability to explain the working mechanism of a task verbally, children are able to nonverbally solve a problem (Alibali & Goldin-Meadow, 1993; Goldin-Meadow et al., 1993). Coding both verbal and nonverbal actions of children can increase our knowledge about learning processes even further.

In the examples presented in the results section, mere coding of the explanations of children was possible since children were interviewed using a strict protocol that focused on the predictions and explanations of children. In classroom settings, children are not always challenged to verbalise their explanations (Duschl & Osborne, 2002). Skill theory is applicable in such situations, but it requires a more elaborate coding scheme and highly trained coders to code all verbal and nonverbal behaviour. This might be a threshold for using skill theory by teachers. One solution can be to train teachers to ask more questions and elicit more explanations from children so that only the explanations need to be scored. Questioning by the teacher is also promoted by Wetzels et al. (2011) and Oliveira (2010) as a means to increase the quality of the interaction between teacher and child during science lessons.

Aside from extending the scheme to other (non-) verbal behaviour, a further classification can be made of the explanations within the same level based on the content. For instance, level 4 on the balance scale is coded when the child understands the role of weight or the role of distance. The same can be found in the floating and sinking task when children argue that one object floats because it is light and another object because it is round. These are same-level but not identical explanations. A further subdivision within levels could be helpful in explaining transition mechanisms within and between levels and tiers (Fischer, 1980; Fischer & Bidell, 2006). A detailed analysis of the content of the explanations can also shed light on the influence of mistakes or misconceptions. Incorrect explanations can be coded at the same level as their correct counterparts since the complexity level is equal, but the misconception might influence the learning process. Schwartz and Fischer (2004) argue that it can lead to a different pathway towards a higher level, but further research has to confirm this.

Although this article is not intended as an empirical study, some issues can be addressed regarding the illustrations presented here. First, the context-specific approach that we use assumes that situational heuristics can be neglected. Second, this context-specific approach states that skills are soft-assembled (Thelen & Smith, 1994), which means that skills are created in interaction with the context. This is in line with the dynamic systems approach of skill theory. The context used in this study was a situation in a separate room, with an unfamiliar adult, one task and a very strict protocol for the adult. This protocol was designed to help children set up and test hypotheses and was very strict for the adult in order to have an almost identical context for all children. However, this context is not one-on-one comparable to class settings or parent-child interactions. The performance of the children presented in this study are reflections of their performance under these settings with this protocol and will be different in other contexts (cf. Meindertsma et al., submitted). In that study, different tasks and different protocols elicited different performance levels in the same group of children. As argued, skill theory can also be applied in these other contexts and can therefore be useful for comparing children's performance in more or less natural situations.

Overall, skill theory is widely applicable and has the potential to enhance the quality of research methods and teachers' educational skills within science education. Future research needs to establish the reliability and ease of use of this coding scheme outside research settings. Furthermore, transition mechanisms between and within levels need to be unravelled as well as the learning pathways following misconceptions.

References

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: what the hands reveal about a child's state of mind. *Cognitive Psychology*, *25*, 468-523. doi:10.1006/cogp.1993.1012

Bassano, D., & Van Geert, P. (2007). Modeling continuity and discontinuity in utterance length: a quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science*, *10*, 588-612. doi:10.1111/j.1467-7687.2006.00629.x

Butts, D. P., Hofman, H. M., & Anderson, M. (1993). Is hands-on experience enough? A study of young children's views of sinking and floating objects. *Journal of Elementary Science Education*, *5*, 50-64. Springer Netherlands. doi:10.1007/BF03170644

Chien, S. C. , Hsiung, C. T., & Chen, S. F. (2009). The development of young children 's science-related concept regarding "floating and sinking." *Asia-Pacific Journal of Research in Early Childhood Education*, *3*, 73-88.

Crowley, K., & Galco, J. (2001). Everyday activity and the development of scientific thinking. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings*. (pp. 393-413). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*, 39-72. Routledge. doi:10.1080/03057260208560187

Eshach, H., & Fried, M. N. (2005). Should Science be Taught in Early Childhood? *Journal of Science Education and Technology*, *14*, 315-336. doi:10.1007/s10956-005-7198-9

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, *87*, 477-531. doi:10.1037/0033-295X.87.6.477

Fischer, K.W., & Bidell, T. R. (2006). Dynamic development of action and thought. In R. M. Lerner & W. Damon (Eds.), (pp. 313-399). Hoboken, NJ US: John Wiley & Sons Inc.

Fischer, K. W., & Dawson, T. L. (2002). A new kind of developmental science: Using models to integrate theory and research. *Monographs of the Society for Research in Child Development*, *67*, 156-167. doi:10.1111/1540-5834.671184

Glauert, E. (2005). Making sense of science in the reception class. *International Journal of Early Years Education*, *13*, 215-233. doi:10.1080/09669760500295862

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*, 279-297. doi:10.1037//0033-295X.100.2.279

Grannot, N., & Parziale, J. (2002). *Microdevelopment: transition processes in development and learning*. Cambridge: Cambridge University Press.

Grannot, N., Fischer, K. W., & Parziale, J. (2002). Bridging to the unknown: A transition mechanism in learning and development. In: N. Grannot & J. Parziale (Eds.), *Microdevelopment: Transition Processes In Development And Learning* (pp. 131-156). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511489709.006

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "Floating and Sinking." *Journal of Educational Psychology*, *98*, 307-326. doi:10.1037/0022-0663.98.2.307

Havu-Nuutinen, S. (2005). Examining young children's conceptual change process in floating and sinking from a social constructivist perspective. *International Journal of Science Education*, *27*, 259-279. doi:10.1080/0950069042000243736

Hsin, C. T., & Wu, H. K. (2011). Using scaffolding strategies to promote young children's scientific understandings of floating and sinking. *Journal of Science Education and Technology*, 1-11. Springer Netherlands. doi:10.1007/s10956-011-9310-7

Jansen, B. R. J., & Van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, *81*, 383-416. doi:10.1006/jecp.2002.2664

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: effect of direct instruction and discovery learning. *Psychological science: a journal of the American Psychological Society / APS*, *15*, 661-7. SAGE Publications. doi:10.1111/j.0956-7976.2004.00737.x

Meindertsma, H. B., Van Dijk, M. W. G., Steenbeek, H.W. & Van Geert, P. L. C. (*submitted*). Assessing preschooler's scientific reasoning during interaction with an adult: what is the optimal context?

Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching*, *47*, 422-453. doi:10.1002/tea.20345

Parziale, J., & Fischer, K. W. (1998). The practical use of skill theory in classrooms. In R. J. Sternberg & W. M. Williams (Eds.), *Intelligence, instruction, and assessment: theory into practice* (pp. 95-110). Mahwah, N.J.: Lawrence Erlbaum Associates.

Philips, S., & Tolmie, A. (2007). Children's performance on and understanding of the balance scale problem: the effects of parental support. *Infant and Child Development*, *16*, 95-117. doi:10.1002/icd.504

Piaget, J. (1930). *The child's conception of physical causality*. New York: Harcourt, Brace & Company.

Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: children learning about balance. *Developmental Psychology*, *40*, 1059-1067.

Rappolt-Schlichtmann, G., Tenenbaum, H. R., Koepke, M. F., & Fischer, K. W. (2007). Transient and robust knowledge: contextual support and the dynamics of children's reasoning about density. *Mind, Brain, and Education*, *1*, 98-108. doi:10.1111/j.1751-228X.2007.00010.x

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, *32*, 102-119. doi:10.1037/0012-1649.32.1.102

Schwartz, M., & Fischer, K. W. (2004). Building general knowledge and skill: cognition and microdevelopment in science learning. In: A. Demetriou & A. Raftopoulos (Eds.), (pp. 157-185). New York, NY US: Cambridge University Press.

Siegler, R. S. (1995). How does change occur: a microgenetic study of number conservation. *Cognitive Psychology*, *28*, 225-73. doi:10.1006/cogp.1995.1006

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: a microgenetic analysis. *Cognitive Psychology*, *36*, 273-310. doi:10.1006/cogp.1998.0686

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481-520. doi:10.1016/0010-0285(76)90016-5

Siegler, R. S., & Svetina, M. (2002). A microgenetic/cross-sectional study of matrix completion: comparing short-term and long-term change. *Child Development*, *73*, 793-809. doi:10.1111/1467-8624.00439

Steenbeek, H. W., Van Geert, P. L. C., & Van Dijk, M. W. G. (2011). Dynamics of children's science. *Netherlands Journal of Psychology*, *66*, 96-109.

Tenenbaum, H. R., Rappolt-Schlichtmann, G., & Zanger, V. V. (2004). Children's learning about water in a museum and in the classroom. *Early Childhood Research Quarterly*, *19*, 40-58. doi:10.1016/j.ecresq.2004.01.008

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Tytler, R., & Peterson, S. (2003). Tracing young children's scientific reasoning. *Research in Science Education*, *33*, 433-465. doi:10.1023/B:RISE.0000005250.04426.67

Tytler, R., & Peterson, S. (2004). From try it and see to strategic exploration: Characterizing young children's scientific reasoning. *Journal of Research in Science Teaching*, *41*, 94-118. doi:10.1002/tea.10126

Van Benthem, J., Dijkgraaf, R., & De Lange, J. (2005). *TalentenKracht*. The Hague.

Van der Steen, S., Steenbeek, H. W., & Van Geert, P. L. C. (*submitted*). EBD students and their understanding of scientific concepts.

Van Dijk, M., & Van Geert, P. (2007). Wobbles, humps and sudden jumps: a case study of continuity, discontinuity and variability in early language development. *Infant and Child Development*, *16*(1), 7-33. doi:10.1002/icd.506.

Van Geert, P. L. C., & Steenbeek, H. W. (2005). The dynamics of scaffolding. *New Ideas in Psychology*, *23*(3), 115-128. doi:10.1016/j.newideapsych.2006.05.003.

Wetzels, A. F. M., Steenbeek, H. W., & Van Geert, P. L. C. (*submitted*). The effect of a coaching program on teachers' well-being in science and technology sessions in primary education.

Yan, Z., & Fischer, K. (2002). Always under construction. *Human Development*, *45*, 141-160. doi:10.1159/000057070

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*, 99-149. doi:10.1006/drev.1999.0497

HEIDI B. MEINDERTSMA
PhD student in developmental psychology at the Faculty of Behavioural and Social Sciences, University of Groningen, the Netherlands. Research topics: short-term dynamics of interaction; scientific reasoning in early childhood.
E-mail: h.meindertsma@rug.nl

MARIJN W.G. VAN DIJK
Lecturer in developmental psychology at the Faculty of Behavioural and Social Sciences, University of Groningen, the Netherlands. Research topics: early interaction and development (feeding, language); dynamics of learning in primary education.
E-mail: m.w.g.van.dijk@rug.nl

HENDERIEN W. STEENBEEK
Lecturer in developmental psychology at the Faculty of Behavioural and Social Sciences, University of Groningen, the Netherlands. Research topics: socially situated learning in primary education; social interaction; dynamic systems model of learning and interaction.
E-mail: h.w.steenbeek@rug.nl

PROF. DR. PAUL L.C. VAN GEERT
Professor of Psychology at the Faculty of Behavioural and Social Sciences, University of Groningen, the Netherlands. Especially in the field of dynamic system models of development. Research topics: elaboration of dynamic systems models of development; language and socio-cognitive development; cognitive change and educational psychology.
E-mail: p.l.c.van.geert@rug.nl

Notes for contributors

The Netherlands Journal of Psychology publishes exclusively in English, UK spelling. Manuscripts must be original and not currently under review or published elsewhere. Articles, Notes and other contributions should not exceed the word limit mentioned elsewhere on the cover without consulting the editor.
If an anonymous review is preferred, the manuscript should include two title pages, one with identifying information and one with the title only. All personal information should be excluded from the body of the text as well. Only manuscripts prepared in this fashion will be subject to blind reviews. Manuscripts are not returned to authors.

The associate editors will assist authors in preparing the manuscript conform the Journal's editorial policy. Because of the broad range of subject matter covered by the Journal, authors are encouraged to supply the names of one or more potential referees. Doing so in no way guarantees that the editors will request reviews from these individuals, although such referees may be called upon in addition to those of the editors' choosing.

Netherlands Journal of Psychology follows the publication style of the American Psychological Association. All manuscripts require an abstract of 100-150 words typed on a separate page and 5-10 key words should be provided. Cited references should be presented in the text by author and date and be collated into a reference list at the end of the article with the following information: author(s), year of publication, title and publishing data. Authors are encouraged to consult the *Publication Manual* of the *American Psychological Association* (5th Edition, see downloadable Style Sheet for quick overview). Authors are requested to restrain from 'overcitation'. Often two or three references will suffice.

Manuscripts not properly prepared will not be considered.
Authors should supply a brief note of about 50 words with their main fields of interest, and including their mailing address, an electronic-mail address, and phone and fax numbers. Authors are responsible for obtaining permission from copyright holders for reproducing any illustrations, tables, figures or lengthy quotations previously published elsewhere. Authors are sent a pdf of the published version.
Copyrights are conform the publisher's guideline (see cover).
Authors must submit manuscripts in electronic formats to:

Rene.vanhezewijk@ou.nl

Address of the editor

Editor of the Netherlands Journal of Psychology
Prof. R. van Hezewijk
Faculty of Psychology OUNL
PO Box 2960
6401 DL Heerlen
The Netherlands