Nursing workload measurment instrument

Validity, reliability and applicability

Mark Haaksma

UMCG, Thoraxcentrum
Rijksuniversiteit Groningen, Faculteit Economie en Bedrijfskunde, Master Operations & Supply Chains

Groningen, 20 April 2009

Studentenbureau UMCG/afstuderen

Universitair Medisch Centrum Groningen

Nursing workload measurement instrument

Validity, Reliability and Applicability

Groningen, 20 april 2009

Auteur	Mark Haaksma
Studentnummer	1582941
Afstudeerscriptie in het kader van	Master Operations & Supply Chains Faculteit Economie en Bedrijfskunde Rijksuniversiteit Groningen
Opdrachtgever	mw. drs. M.W. Posthumus mw. L.J.A. Redelaar Thoraxcentrum, UMCG
Begeleiders	dr. M.J. Land dr. M.J. Oosterhuis Faculteit Economie en Bedrijfskunde Rijksuniversiteit Groningen
Begeleider UMCG	dr. J. Pols Studentenbureau UMCG / afstuderen

ISBN [Klik hier en typ ISBN-nummer]
NUR 801
Trefw Nursing Workload Measurement Instrument

Omslag: Wenckebach Instituut, Universitair Medisch Centrum Groningen

Preface

In May 2008 I started this graduation project, researching a nursing workload measurement instrument for the Thorax centre of the University Medical Hospital in Groningen. During this individual project I learned a lot regarding applying my operations knowledge to a hospital environment and managing my own project. Although this was an individual project I could not have done it without the help of the following people:

- The supervisors of the university, dr. M.J. Oosterhuis and dr. M.J. Land.
- The management of the Thorax Centre and company supervisors, mrs. drs. M.W. Posthumus and mrs. L.J.A. Redelaar.
- The people of the Student Agency / graduation, who provided me with any help necessary, dr. J. Pols and mrs. A. Muurman.

I hope this master thesis will be interesting, knowledgeable and enjoyable to read.

Mark Haaksma

Table of contents

Management summary

The management of the thorax centre of the UMCG needs a good insight in personnel, equipment and materials needed to provide the necessary quality of care. It is hard to determine the number of nurses that are needed in future time periods. The management decided to develop a workload measurement instrument, in which every action a nurse performs on a specific patient is documented, aiming to get more insight in the personnel needed.

After two years of using the instrument a group of students from the University of Groningen were asked to evaluate the instrument. They suggested that the validity and the reliability of the measurement instrument are questionable. This led to the following research question: What are the validity, reliability and the applicability of the measurement instrument used by the thorax centre of the UMCG?
In order to answer this question different research methods were applied. Firstly, a literature review was performed leading to the requirements which an applicable measurement instrument needs to meet. The nursing workload measurement instrument of the UMCG is in this thesis tested against each of these requirements. After the literature review the data needed to answer these sub questions was obtained using observations, interviews and the historic data of the measurement instrument. Six of the nurses of the thorax centre were observed during a dayshift and interviewed after that dayshift. During these observations every activity of the nurse was documented and timed. The gathered data was later compared to the historical data collected with the current nursing workload measurement instrument.

Analysis of the collected data leads to the conclusion that the nursing workload measurement instrument is not reaching its full potential. The activities listed in the measurement instrument are not fully covering the definition of nursing work and the operation time of each of these items are not matching the actual times needed by the nurses. Therefore, the current measurement instrument is not supporting the decision making process of the management. However, the measurement instrument has potential. When the measurement instrument is improved the management will have a useful tool to support their budget requests and get insight in the nursing work on their wards.

1 Introduction

The sections 1.1 and 1.2 of this chapter introduce the thorax centre, a department of the university hospital of Groningen (UMCG), on which this research focuses. In the sections 1.3 and 1.4 the research design and the methodology of this research project are clarified.

1.1 The University Hospital of Groningen

The UMCG is the only university medical centre in the northern part of the Netherlands, and therefore the final point of referral for many patients. Patients go to the UMCG for basic care as well as highly specialist top clinical and top reference care, such as organ transplants, complex neuro-surgery, neonatology, clinical genetics, In Vitro Fertilization (IVF), pediatric oncology, renal dialysis and traumatology. All medical specialties are represented, as well as education programs for all medical disciplines.

The more than 8500 employees and 1300 beds make the UMCG one of the largest hospitals in The Netherlands. The UMCG is one of the largest employers in Groningen and a steadily expanding organization. Each year the number of employees increases by 150. These employees are necessary in order to maintain the status of university medical centre. The UMCG's philosophy with regard to patient care extends beyond the hospital's walls. The various aspects of care should be finely tuned to one another, and the UMCG takes this responsibility seriously. The UMCG cooperates closely with various care institutions and nursing homes.

The UMCG is working in various ways to improve its quality of care and service. Each department makes a clear assessment of the level of its care and service, so as to be able to recognize possible areas for improvement and to take the proper measures. In this, the opinion of the patients is paramount. They have experienced care at the hospital, and are likely to have ideas for improvements. They are a vital source of information.

1.2 The thorax centre

The department of the UMCG on which this research focuses is the thorax centre. The thorax centre has two wards, namely B1VA and B2VA, with each thirty two beds. The medical specialties Cardio Thoracic Surgery and Cardiology together form the thorax centre in the UMCG.

Cardiology is the specialty that deals with the diagnosis and treatment of heart disease. For example, the cardiologist treats patients with diseases caused by constrictions in the coronary arteries. The specialist also devotes attention to risk factors for the emergence of heart disease, such as high blood pressure, high cholesterol and unhealthy living habits. The cardiologist often performs functions as heart film studies, echocardiography, blood tests and hart catheterization.

Cardio Thoracic surgery is the specialty that focuses on the surgical treatment of heart and lung diseases. Thoracic Surgeons perform under more bypass operations, operations of congenital heart defects, operations to treat heart arrhythmia, lung surgery and operations in which one or more heart valves are repaired or replaced.

1.3 Research design

1.3.1 Initial motive

The management of the thorax centre of the UMCG needed a good insight in personnel, equipment and materials needed to provide the necessary quality of care. It is challenging to determine the number of nurses that are needed the next time period. The management decided

that a workload measurement instrument, in which every action a nurse performs on a specific patient is documented, could be helpful in getting more insight in the personnel needed. A measurement instrument also provided the management with a strong case when they struggle to get more personnel. The management decided not to use an existing measurement instrument, but developed an instrument specific for the thorax centre.

The management takes part in a national project group with representatives of other thorax centers from around the country. After developing and implementing the measurement instrument they presented it to the representatives of the other hospitals. The other hospitals immediately showed their interest in the instrument and some even want to implement it in their centre.

1.3.2 Problem statement

Before the management lets other thorax centers use the instrument they want to be sure it works properly. After two years of using the system a group of students from the University of Groningen were asked to evaluate the instrument. They concluded that the validity and the reliability of the instrument are questionable. Also they gave the advice to research the further possibilities of the instrument and the stored data. Furthermore, the management of the thorax centre would like to simplify the current measurement instrument. The objective of this research is to determine if the used nursing workload measurement instrument, currently used by the thorax centre, has enough quality to be used in the decision-making process of the management and ultimately to be implemented in other hospitals.

For this thesis project it led to the following research question: *To what extent is the measurement instrument used by the thorax centre of the UMCG valid, reliable and applicable and how can it be improved?*

1.3.3 Sub questions

To achieve this objective the research question is divided into three topics with each one or more questions:

Applicability:

1. What requirements need to be met to make an applicable nursing workload measurement instrument?
2. Which choices are made during the development of the current nursing workload measuring instrument?
3. How does the management team use the information provided by the nursing workload measurement instrument in the current situation?
4. What are the possibilities to simplify the nursing workload measurement instrument?
5. What are the possibilities to further improve the applicability of the nursing workload measurement instrument?

Validity:

6. To what extent is the nursing workload measurement instrument valid considering the content validity of the instrument?
7. To what extent is the nursing workload measurement instrument valid considering the construct validity of the instrument?
8. What are the possibilities to improve the validity of the nursing workload measurement instrument?

Reliability:

9. To what extent is the nursing workload measurement instrument reliable considering the test re-test reliability of the instrument?
10. To what extent is the nursing workload measurement instrument reliable considering the inter-observer reproducibility of the instrument?
11. What are the possibilities to improve the reliability of the nursing workload measurement instrument?

1.4 Methodology

To answer the research question the above formulated sub questions need to be answered. In this section the different research methods applied to answer these questions are clarified. In this research most of the answers are based on observations, besides interviews and a literature review.

Observations
During this research six nurses are observed during their dayshift. These observations form the basis for answering the sub questions regarding the validity of the measurement instrument, which is an important part of this research. For this research an observation scheme (appendix A) and a code scheme (appendix B) are developed, based on the existing list of items in the current measurement instrument. In the six days of observing the aim is to collect sufficient data to gain insight in the activities the nursing work of the thorax centre consists of and to determine the time needed for each of these activities. The observation period of six days is cut into two series of three days, to have the possibility to make intermediate changes if one of the schemes doesn't work correctly.

Interviews
After being observed during their dayshift an interview is conducted with each of the nurses. The data collected by interviewing the nurses is used when answering the sub questions regarding the reliability and applicability of the nursing workload measuring instrument. Insight in the decision making process of the nurses during the use of the measurement instrument, gained with interviews as well as observations, forms the basis for the answering of the reliability sub questions. Furthermore, by interviewing the nurses insight is gained in the thoughts of the staff on the applicability and how to improve the applicability of the measurement instrument. The interviews are conducted semi-structured as can be seen in appendix C.

Literature review
The first part of the research consists of a literature review. The purpose of this literature review is twofold. Firstly, the four points a measurement instrument needs to comply with are concluded based on this review. Based on these four points conclusions are made and the possibilities to improve the measurement instrument are advised. Secondly, the sub questions regarding validity and reliability are based on assessments made in the literature review.

Historical data
The data collected with the nursing workload measurement instrument since 2004 is also used in this research. This historical data is used to answer the questions regarding the reliability and the applicability of the measurement instrument. With the analysis of the historical data a good insight is gained in the usage of the measurement instrument over the past five years and the output the measurement instrument has produced during this period.

Sub questions	Chapter	Obser-vation	Inter-views	Literature review	Historical data
1	2.7			X	
2	4.1		X		
3	3		X		X
4	4.2	X			
5	6			X	
6	4.3	X			
7	4.2	X			
8	6				
9	4.4	X	X		X
10	4.4	X	X		X
11	6				

Table 1: The research methods applied for every sub question and in which chapter the collected data and the answer for the sub question is presented

1.5 Boundaries and Scope

In this section the boundaries and scope of this research are described. Both the boundaries and scope of the research are important to clarify during research to maintain consistency and understanding of every person involved in the research.

1.5.1 Boundaries

This research focuses on the nursing workload measurement instrument used by the two nursing wards of the Thorax centre of the UMCG. The aim of the research is to make the measurement instrument valid and reliable for these wards, therefore all the observations are done on these wards. Other wards of the thorax centre are left out of this research.

1.5.2 Scope

As mentioned this research aims to get insight in the validity, reliability and applicability of the nursing workload measurement instrument currently used by the thorax centre. The result of this research will be an advice which can be followed by an IT professional to make the measurement instrument valid and reliable. There will be no actual changes done to the current measurement instrument during this research.

2 Theoretical framework

In order to answer the research question it is important to review the literature that forms the basis for the research. In this chapter the literature regarding nursing workload measurement instruments is reviewed, starting with a definition of the terms nursing work and nursing workload. After the definitions are stated the review focuses on the measuring the nursing workload with computerized instruments and the difficulties with these instruments. In section 2.6 the literature regarding the validity and the reliability of nursing workload measurement are described. This part of the review is used to divide the research question into the sub questions described in section 1.3. Finally, the review leads to three requirements that need to be met by a nursing workload measurement instrument in order to be applicable.

2.1 Nursing

Nursing is commonly defined in functional terms, mostly in terms of the actions, work or activities carried out by the nurse. Henderson (1966, p. 15) defined nursing as follows 'the unique function of the nurse is to assist the individual, sick or well, in the performance of those activities contributing to health or its recovery (or to a peaceful death) that he would perform unaided if he had the necessary strength, will or knowledge and to do this in such a way as to help him gain independence as rapidly as possible'. A more recent definition of nursing by the American Nurses' Association (ANA) again defines nursing in functional terms. ANA (2003) describes nursing as 'the protection, promotion and optimisation of health and abilities, prevention of illness and injury, alleviation of suffering through the diagnosis and treatment of human response, and advocacy in the care of individuals, families, communities and populations'.

These definitions describe nursing work in terms of what the nurse does for the patient and the function of the nurse is dependent on the area of health care and the physical environment in which the nurse is working. However, in modern society, nursing work also includes non-patient specific activity, e.g. nurse education, management of healthcare systems and administration. The International Council of Nurses (ICN) definition of nursing takes a more global perspective in stating that 'nursing encompasses autonomous and collaborative care of individuals of all ages, families, groups and communities, sick or well and in all settings. Nursing is in this research defined as the promotion of health, prevention of illness, and the care of ill, disabled and dying people.

2.2 Nursing workload

There is no one common definition of nursing workload in the literature. Caplan & Jones (1975) defined workload in terms of the amount of performance required to carry out any job. Arthur & James (1994) defined nursing workload as the 'volume and level of nursing work'. Considering these definitions of nursing workload, it is appropriate to refer back to the ICN definition of nursing work. This recognizes the fact that nursing work comprises both the work that the nurse carries out on behalf of the patient and non-patient-related work, e.g. education of other nurses, organizational and administrative work.

Needham (1997) and other authors have gone further in their definition of nursing workload to consider it in terms of the time taken to carry out 'direct' and 'indirect' care as well other activities, including ward and organization management. Prescott et al. (1991) discuss both direct and indirect care, where direct care accounts for all nursing activities carried out in the presence of the patient and/or family and indirect care accounts for any work carried out away from but on behalf of a specific patient. O'Brien et al. (2002) support this thinking, highlighting that when

measuring nursing workload it is essential to distinguish between 'patient' and 'non-patient' related work, and then to link this distinction to an ideal skill mix.

Integration of the ICN definition of nursing work with Needham's definition of nursing workload, and Sovie and Smith's (1986) and O'Brien et al. (2002) idea that direct and indirect patient care activities differ from other nursing activities that are related to organizational management activity rather than patient care, gives us an overview of how these nursing work-related concepts come together to form an initial model of the conceptual makeup of nursing workload (see Figure 1).

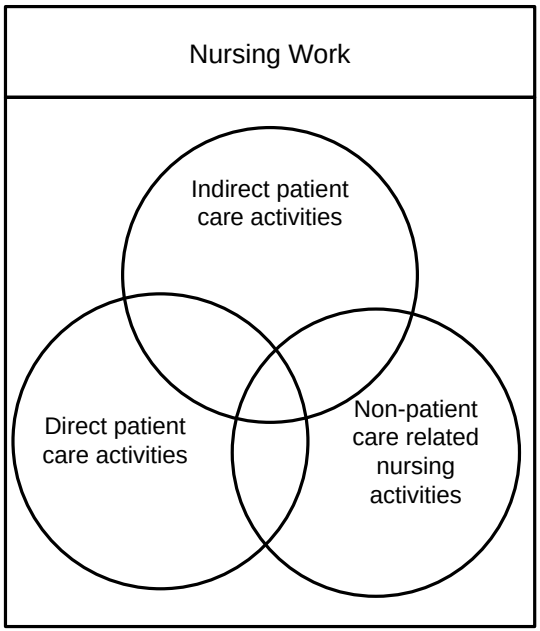

Figure 1: The make-up of nursing work (based on the International Council of Nurses definition of Nursing).

2.3 Measuring nursing workload

Bridel (1993) suggests that the measurement of nurses' workload allows the estimation of staffing requirement, education and cost and the matching of the actual patient need with the actual nursing hours. A workload measurement system needs to reflect consumer care needs on each shift and include time allowance for acute and unexpected interventions (Albie-Gibbons, 1986). A major variable in staffing formulas is the amount of time nurses spend directly (face-to-face) caring for patients but the appropriate direct care percentage has never been established. Levels around 35% are often used (O'Brien, 1986) although Carr-Hill & Jenkins-Clarke (1995) noted that direct care time greater than 50% was associated with higher-quality care. Confusion and uncertainty were increased when recent data indicated that nursing direct care time was falling yearly without a corresponding reduction in nursing quality (Hurst, 2003; McKenna, 1995).

Besides the disagreement regarding the percentage of nursing workload appropriate for direct care there are many difficulties inherent in measuring nursing workload. The physical aspects of nursing are easy to observe and measure, yet the fundamental but critical aspects, such as decision-making and caring, are difficult to quantify (Brady et al., 2007). Nursing is labour intensive and service driven and is difficult to measure because the inputs, outputs or outcomes of nursing care are not always tangible or sufficiently concrete to facilitate measurement by a third party (Grandfield, 1992). Also, nurses do many tasks simultaneously in the process known as multitasking; for example, educating a patient while doing their dressing (Endacott & Chellel, 1996). Furthermore, nursing intensity of work is influenced by multiple factors, including the nurse's characteristics such as age, skill, experience and education and the complexity of the team and work environment (O'Brien-Pallas, 1997, 2001).

The volume of data and paper generated would tend to negate the use of a manual system. Therefore, the choice of the UMCG in 2004 to design a computerized measurement instrument is obvious. However, several problems may occur when using computerized workload measurement systems. In their document Making Best Use of Ward Nursing Resources the audit commission (1991) reports that systems in place have often produced a mass of data but little improvement. This is echoed by Coles & Jenkins (1992) who write that information produced must be relevant and help with the decision making. Therefore, an important aspect of the nursing workload measurement instrument must be to produce relevant information.

2.4 Nursing workload measurement instruments

Awareness of nursing workload measurement and skill mix has risen during the last two decades as the financial implications of healthcare have become apparent. Walker and Whynes (1990) argued that the proportionately high cost of the nursing budget impinged heavily on the cost of treatment costs. Nurses, as the largest staff group, make up the highest labour cost while labour consumes a large proportion of the healthcare budget. Development of nursing workload measurement instruments has been influenced by changes in health funding and service provision. Managers and funders of health care now require knowing the hours and skill level of nursing needed to produce consumer outcomes.

In an earlier report, Wright-Warren (1986) concluded that ward management needed supporting information to determine the skill mix, level and efficient utilization of staff. The goal of workload measurement is that nursing costs can be reduced or justified if nurses were deployed efficiently. The Resource Management Initiative (DHSS 1986) proposed that information technology was the tool to provide the supporting information. It also encouraged the use of workload assessment methods to determine skill mix requirements. Workload measurement is an attempt to predict the nursing time and skills required to provide nursing care. The Strategy for Nursing (1989) also promoted the use of computers to help quantify nursing workload and 'workload measurement of nursing care' was identified by the NHS Management Executive (1989) as one of seven information requirements for ward managers. These powerful influences (quantifying nursing input and efficient deployment of nurses) have led to the adoption of a variety of nursing workload measurement methods (Hughes, 1999).

As described by Hughes (1999) there are two main methods of measuring workload: activity based and dependency based. These are also referred to by Edwardson and Giovannetti (1994) as 'nursing task document' and 'critical indicators of care' approaches. Activity methods involve an assessment of the patient for the activities involved in providing the nursing care required. Each activity has, through observation, been allocated a time to carry it out. The sum of a patient's required nursing activity times should supply the total time required to care for that patient in that shift or day.

Not requiring such detailed information, dependency methods require a simpler assessment. One method uses a simple classification or critical indicators tool that places the patient in one of four categories. Numbered one to four, the first category represents patients who care mostly for themselves while category four patients require high input of care. Each category has been assigned a nursing time. In this way the total time required for nursing care can be identified for a 24-hour period from the number of patients in each category. Important for workforce planners is the best way of classifying patients for estimating nursing workload. There are a variety of approaches in the literature. Cambell et al. (1997), Needleman et al. (2002), for example, in their North American studies used casemix based on diagnosis related groups (DRG's) to categorize patients for nurse staffing purposes.

In the UK, on the other hand, patients are usually classified from least to most dependent on nurses for their activities of living (ADL) needs (Gibbs et al., 1991). Both approaches have strengths and weaknesses. On the one hand DRG methods hit a stumbling block when there is co-morbidity making a decision about which illness or treatment takes priority for workload assessment purposes difficult. The ADL method, on the other hand, is prone to inflated scores by nurse assessors who know that higher acuities indicate that wards may be short staffed (Hurst, 2003; Whitney and Killien, 1987).

A third approach is the statistical method. This is more a nursing requirement predictor tool than a workload assessment method. Statistical methods base their predictions on previous information from the clinical area itself. From each shift, information is collected on the number of significant patient events such as admission, discharge and patient-related procedures. The nurse in charge also assesses the shift for any unsafe incidents and how much of the work was achieved. In conjunction with the duty roster, regression analysis is carried out to identify appropriate grade mixes for different permutations of patient events.

All three of these methods lend themselves to computerization and, on the surface, this provides a legitimate context for their acceptance. For a method to be easily transferred to a computer to provide sensible, quantitative data then it would seem that it must be based on a sound, empirical framework. However, several studies that have been carried out to test this legitimacy have not found an empirical framework (Hughes, 1999). Many workload measurement tools document direct nursing care (patient-centered nursing care activities) but fail to document the indirect nursing care (Brady et al., 2007). Considering this the nursing workload measurement instrument needs to cover all three elements of nursing work shown in figure 1.

2.5 Difficulties with nursing workload measurement instruments

In this chapter the advantages of nursing workload measurement instruments are described, but the literature also describes the known difficulties with nursing workload measurement instruments. Seago (2002) argues that without the ability to control input (patient admissions) to the unit, there is no way to predict the workload.

Criticisms of workload measurement include the complexity of nursing care given or the quality of the care (Benefield, 1996). Even with the most sophisticated organizational information systems to record face-to-face contact with clients, it is difficult to quantify how much time was spent on the provision of different types of care, such as education and advice or administration of treatment (Freeman et al. 1999). In a survey of patients within different categories of a dependency method Stillwell and Hawley (1993) found that the actual nursing care times were extremely varied between patients in the same category.

It is suggested (Greenhalgh et al. 1990) that patients are becoming more homogenous in their dependency levels as patient stays become shorter. If this is so then the dependency workload methods have less use; if most patients are assigned to only one or two categories then the associated workload requirements for these patients start to merge. Under these circumstances the number of patients would become the workload indicator and assessment methods would provide little in the way of extra, useful information.

A criticism of activity-based systems is the focus on the tasks performed with insufficient measure of the less tangible aspects of the work. The time taken to perform nursing activities is often governed by the individual needs of the patient (Hughes 1999).

Arthur and James (1994) suggested that 'a perfect tool for measuring nursing workload is unlikely to exist.' If the workload assessment methods are, at present, unreliable then their use needs to be questioned. If the purpose of nursing is to provide an agreed quality of care, and managers want to provide this as efficiently as possible, then the two measures need to be the quality of nursing care and nursing costs. Hughes (1999) states that using workload measures to predict staffing has been demonstrated to be flawed but may be useful to indicate changes. In conclusion, an applicable nursing workload measurement instrument needs to focus on showing changes in nursing workload and trends in the nursing work instead of trying to predict the exact staffing needed.

2.6 Validity and reliability

The importance of establishing validity and reliability for nursing workload measurement instruments is emphasised in the literature (Hernandez & O'Brien-Pallas, 1996), but many studies report minimal validity and reliability or omit to report validity or reliability, and lack applicability to specialist settings (Davidhizar, 1998). Several authors have commented on the lack of rigorous studies conducted into proprietary systems (Edwardson & Giovanetti, 1994; Hernandez & O'Brien-Pallas, 1996; Malloch, & Conovaloff, 1999), and others report a lack of predictive validity in existing instruments (Carr-Hill & Jenkins-Clarke, 1995). Actual times spent in care were greater than the times predicted by the instruments used. Hughes (1999) concluded that there is little reliability to existing instruments, either because of poor inter-rater reliability, or problems inherent in the methods of measurement.

2.6.1 Validity

Valid measurement is achieved when scores (including the results of qualitative classification) meaningfully capture the ideas contained in the corresponding concept (Adcock & Collier, 2001). In many arenas, validity is the most important attribute (Lohr et al., 1996). Lohr et al. (1996) developed scientific review criteria for evaluating instruments based on the classis definition of the Scientific Advisory Committee (SAC) of the Medical Outcomes Trust namely, the degree to which an instrument measures what it purports to measure. There are three ways of accumulating evidence for the validity of an instrument:

Content-related
Content-related evidence shows that the content domain of an instrument is appropriate relative to its intended use. Methods commonly used to obtain evidence about content-related validity include the use of lay and expert panel judgments of the clarity, comprehensiveness, and redundancy of items and scales of an instrument. The process of establishing content validity usually starts with a definition of the concept that the investigator is attempting to measure. A second step is a literature search to see how this concept is represented in the literature (Morgan et al., 2001).

Construct-related
Construct-related evidence supports a proposed interpretation of scores on the instrument based on theoretical implications associated with the construct 'Nursing work'. Common methods to obtain construct-related validity include an examination of the logical relations that should exist with other measures and/or patterns of scores across groups of individuals. This validation procedure usually involves establishing a relationship between the instrument and a measurable external or outside criterion (Morgan et al., 2001). The test developer hopes that the less expensive or time-consuming measure will provide very similar information and, thus, a high correlation with the criterion (Cronbach, 1990).

When applying construct validity to an instrument, there is a requirement that the construct which the instrument is measuring is guided by an underlying theory. Often, especially in applied disciplines, there is little underlying

theory to support the construct. Construct validation is a process (relatively slow process) in which the investigator conducts studies in an attempt to demonstrate that the instrument is measuring a construct (Morgan et al., 2001).

Criterion-related
Criterion-related evidence shows the extent to which scores of the instrument are related to a criterion measure. Criterion measures are measures of the target construct that are widely accepted valid measures of that construct. In the area of health status assessment, criterion-related validity is rarely tested because of the absence of widely accepted criterion measures. Cronbach (1990) pointed out that sometimes a test is used for a long time before a theory is developed around it. The aim of this research is to establish the validity and reliability of the current measurement instrument and not to relate the output of the measurement instrument to other measures or theories, therefore criterion validity should be excluded from the research.

Even though an instrument may be consistent (high reliability) it may not be valid (Morgan et al., 2001). For example, one could construct a device for measuring weight. Suppose that we know an object weighs 12 pounds, but the device consistently records 13 pounds. The device would be reliable but not valid for assessing weight because it does not measure what it is supposed to measure.

2.6.2 Reliability

The reliability of a measurement instrument refers to the extent to which the same results are achieved through repeated use of the measurement instrument (Polit & Hungler, 2003). If one uses a standardized instrument, reliability indices should have been established and published in an instrument manual, often referred to in the original journal publication. The method section of a research article also should provide evidence about the reliability of the instrument (Gliner et al., 2001). The two approaches for examining test reliability recommended Lohr et al. (1996) are coefficient (Y (Cronbach's alpha) and alternative form correlations. Because the latter approach is seldom used in health status assessment, the coefficient 0: can be considered the most relevant approach to reliability estimation.

A second definition of reliability is reproducibility or stability of an instrument over time (test-retest) and inter-rater agreement at one point in time. The two definitions are largely independent of one another. Because of the nursing workload measurement instrument is just documenting the performed actions on patients the emphasis of this research is on the reproducibility of the measurement instrument.

Test-retest reproducibility
Test-retest reproducibility as described by Lohr et al. (1996) is the degree to which an instrument yields stable scores over time among respondents who are assumed not to have changed on the domains being assessed. The influence of test administration on the second administration may overestimate reliability. Conversely, variations in health, learning, reaction, or regression to the mean may yield test-retest data underestimating reliability. Despite these cautions, information on test-retest reproducibility data is important for the evaluation of the instrument.

Inter-observer (interviewer) reproducibility
Equivalence or inter-rater reliability refers to the extent to which different nurses using the same workload measurement instrument to measure the same individual at the same time will derive consistent results and this may be determined through evaluation of the percentage agreement between nurses in relation to data generated (Hernandez & O'Brien-Pallas 1996).

2.6.3 Operational definitions

In order to work with these definitions in the research it is important to establish operational definitions. As mentioned in the previous section, criterion validity will not

be included in the research. The following definitions are used for the other criteria:

Content validity

Content validity is interpreted as the degree of matching between the operation times per item used in the measurement instrument and the real operation times.

Construct validity

In this thesis construct validity is interpreted as the extent to which the total list of items in the existing measurement instrument matches the conceptual model of nursing work shown in figure 1.

Test re-test reliability

The operational definition of test re-test reliability is the degree of matching between scores when one nurse uses the measurement instrument multiple times.

Inter-observer reproducibility

Inter-observer reproducibility is used as the degree of matching between scores when different nurses score the same patient.

2.7 Conclusion

In this chapter the requirements for an applicable nursing workload measurement instrument are described. These requirements focus on the validity and reliability of the measurement instrument.

Firstly, the measurement instrument needs cover all the three aspects of nursing work. When not every part of the nursing work is documented in the measurement instrument the output of the measurement instrument will never give a complete insight in the nursing workload.

Secondly, the operation times used in the measurement instrument need to match the actual operation times. Without the match between the times used and the actual times the output of the measurement instrument does not show a nursing workload that represents the nursing workload on the ward. The total nursing workload in the output has no meaning and can only be used to see a trend in the amount of activities that are performed by the nurses.

Finally, to obtain useful output all of the nurses need to use the measurement instrument in the same manner. If every nurse has his or her own interpretation of the way to use the measurement instrument the output will change with the formation of the nursing team. However, if every nurse uses the measurement instrument in the same manner comparisons can be made by the management between periods of time.

In short, the nursing workload measurement instrument of the thorax centre needs to comply with the following points:

- The measurement instrument needs to cover all the three aspects of nursing work shown in the conceptual model of figure 1.
- The time per activity needs to reconcile with the time a nurse spends in practice on that particular activity.
- The users of the measurement instrument need all to use the measurement instrument in the same manner.

3 The current measurement instrument

In this section the current measurement instrument is described. The measurement instrument consists of three screens, namely the ward overview, the actions overview and the output. As mentioned the measurement instrument is in use since 2004 and is used every day by the nurses. In this chapter the three screens of the measurement instrument are described and explained.

The measurement instrument is used at the end or near the end of every dayshift. Every nurse in the dayshift is obliged to score the actions they performed on a specific patient. Usually a nurse is responsible for four to six patients during the dayshift, every one of these patients need to be scored separately in the measurement instrument.

3.1 The ward overview

To start with scoring a patient a nurse opens the ward overview (figure 2) in which he or she can find the patients that need to be scored. In this screen the lay-out of the ward is visible with every bed on the ward. A right-mouse click on a bed opens the measurement instrument for that specific patient (figure 3). After a patient is scored the bed is marked with the sign 'zorgzwaarte ingevuld' and the measurement instrument can not be opened for these patients on this day.

The screen of figure 2 is the starting point of multiple management tasks as well as the starting point of the measurement instrument.

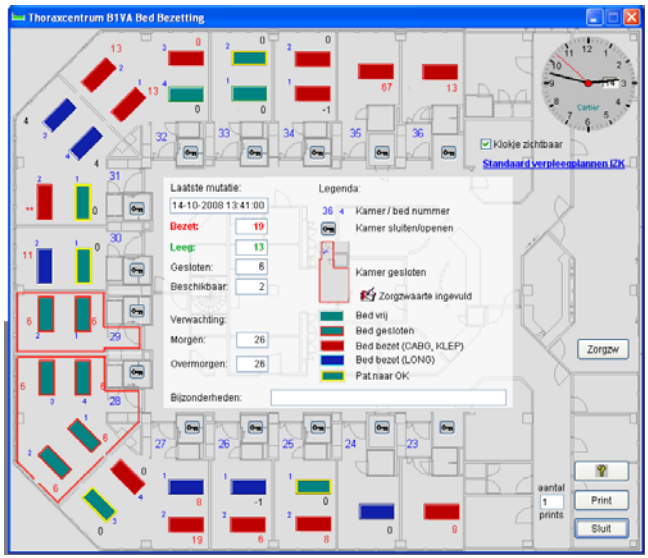

Figure 2: The ward overview of the measurement instrument

3.2 The actions overview

After the right-mouse click on a bed the program shows the actual measurement instrument (figure 3). In this screen the nurse can select every action he or she has performed on the specific patient and save this score. A score can only be saved once a day for every patient, after a save the bed is marked in the ward overview and the measurement instrument cannot be opened for that specific patient.

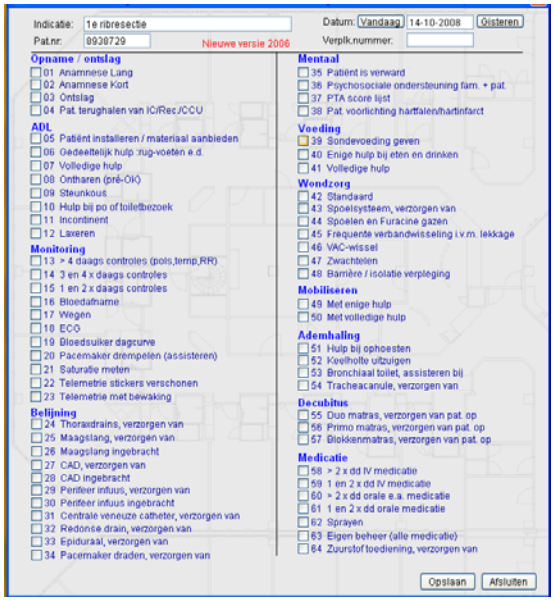

Figure 3: the measurement instrument

In this screen a nurse can choose if an action is performed on the patient or not. If an action is performed several times on the same patient this is not scored in the measurement instrument. In the topics monitoring and medication the times an action is performed can be chosen, because the amount of times the action is performed is mentioned in the description. Beside the actions performed on the patient the pathology of the patient can be indicated in the text field 'indicatie'.

3.3 The output

The third and final screen of the measurement instrument is the output screen (figure 4). This screen is shown when the button 'Zorgzw' is clicked in the ward overview. In the output diagram a specific pathology (ziektebeeld or item) can be choosen to see how many patients with that specific pathology were treated the past month and the nursing workload this resulted in.

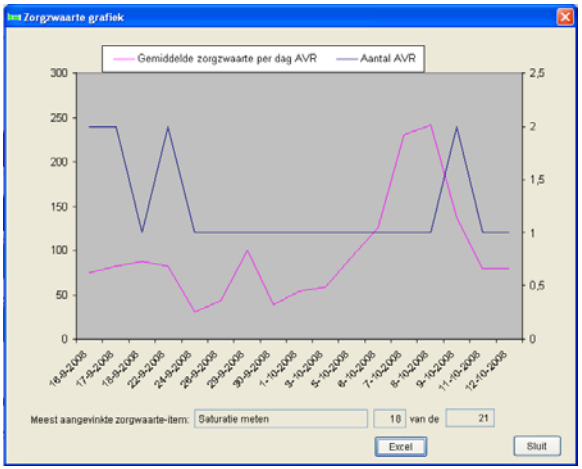

Figure 4: the output of the measurement instrument.

Besides the main diagram the screen also shows the most checked item for this pathology and the button 'excel'. When this button is clicked the data, on which the diagram is based, will be loaded in to Microsoft Excel and additional analysis can be made.

So far in the thesis the methodology, the literature review and the current nursing workload measurement instrument are described. Literature review resulted in the requirements which the nursing workload measurement instrument should meet to make it applicable. In the following chapter for each of these requirements will be established if the current nursing workload measurement instrument meets this requirement using the data collection methods described in sections 1.4.

4 The results

The thorax centre has implemented a measurement instrument on these two wards to improve the quality of care and service. This research focuses on determining whether this measurement instrument is applicable, valid and reliable.

In the following chapter the collected data are elaborated in four different sections, each ending with the answering of the sub questions the section regards. The first section shows what could be found about the development process of the measurement instrument. The second section focuses on the items which are used and which are possibly missing in the measurement instrument. The third section describes the results regarding the value in minutes every item in the measurement instrument has. The fourth and final section focuses on the users and the use of the measurement instrument.

4.1 The development process

When researching the development process no specific documentation of this process and the choices made could be found. Only a PowerPoint presentation in which the measurement instrument was introduced could be recovered. Besides the presentation the only one who could answer some questions was a co-developer, who still works at the UMCG. The following steps of the original development process were recovered from the presentation.

1. An inquiry is held by the head nurse, who worked at the department at that time, in which every nurse could score every item of the instrument. The score they gave represented the number of minutes an action would take.
2. From these scores the median is calculated for each item.
3. In a latter phase of the development process the scores are updated to exclude the scores thought to be too high for a particular item.

This overview of the development process shows that it is not clear in which way every score is calculated. For some items the median is taken and others this median is updated 'to a more convenient' score. In addition it is not clear where the list of items needed for this particular measurement instrument is derived from. The list of items is developed in consultation with the former head nurses, but again the process of developing and the choices made are not documented. However, it is clear that the choice is made to only add direct patient care activities to the measurement instrument, but why this choice is made is not made clear in any documentation.

4.1.1 Conclusion

In order to let other thorax centres implement the nursing workload measurement instrument it is necessary to write design documents. Firstly, without these documents it is impossible for higher management to use the output of the measurement in the decision making process, because they cannot judge how the output is created. Secondly, the lack of documentation makes it very hard to design changes to improve the alignment of the measurement instrument with their nursing work. When changes in the nursing work occur the whole measurement instrument needs to be researched to know if it still concurs with the actual content of the nursing work.

However, the management of the thorax centre of the UMCG knows how the output of the measurement instrument needs to be interpreted. Without design document it is hard for a third party to interpret the output of the measurement instrument and to trust the creation of the information provided. When the management of the thorax centre wants to use the output of the measurement

instrument in budget meetings with the upper management the representatives of the upper management also need to understand the provided information.

4.2 The items in the measurement instrument

This section answers the sub questions regarding the validity aspect of the measurement instrument. As mentioned in the section 2.7 an applicable measurement instrument needs to cover all the three aspects of nursing work and the operation times used in the measurement instrument need to match the actual operation times.

Besides answering the validity sub questions, a proposition is done in this section to remove some of the items currently in the measurement instrument to simplify the instrument.

4.2.1 Conceptual model

In this section the items of the measurement instrument are compared to the model of nursing work shown in figure 1. The model shows that nursing work consists of three elements, namely direct patient care activities, indirect nursing activities and non-patient care related nursing activities

The first thing noticed when comparing the model to the measurement instrument is the lack of indirect nursing activities and non-patient care related nursing activities. Almost all of the items are direct patient care activities. Some indirect nursing activities are implemented in the measurement instrument but most are neglected. Non-patient care related nursing activities are intentionally neglected as a whole when designing the measurement instrument. However, such a choice can be made during the design process, but the time needed for these activities cannot be kept out of the final score when it should reflect the total workload.

Secondly a few of the items are not even activities, but regard the status of a patient. For example the item 'Barrier/isolation nursing' (Dutch: 'Barriere/isolatie verpleging') can be ticked meaning that extra measures need to be taken every time a nurse wants to perform activities on the patient. To calculate the nursing work the time of every activity performed on the patient should be extended with a few minutes. In the current measurement instrument this status is used as an activity and therefore it contributes a single amount of minutes to the total score of the patient not regarding how many activities are performed on the patient.

4.2.2 Rarely used items

In this section a proposition is done to remove some of the items currently in the measurement instrument based on the historical data collected by the instrument itself.

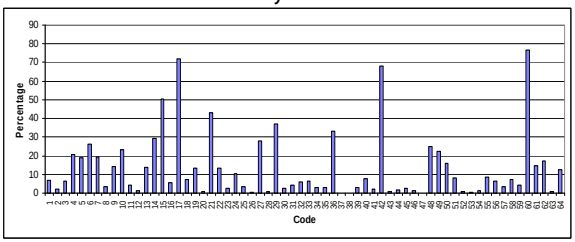

Figure 4: The frequency of codes used in department B1VA.

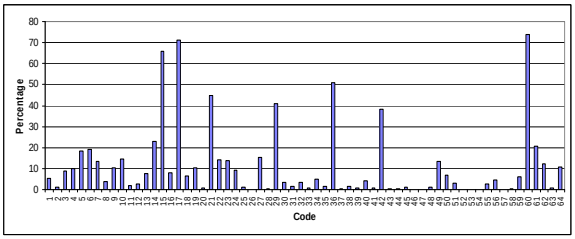

Figure 5: The frequency of codes used in department B2VA.

The figures 4 and 5 show per department the percentage of times a code is used in scoring a patient since the implementation of the measurement instrument in 2004. In total there are 18433 rows in the database for

department B1VA and 16373 rows for B2VA. Every row in the database represents the score for one patient on one day.
The figures clearly show that there are a number of items in the measurement instrument that are almost never used when scoring a patient. Several codes are used in less then one percent of the scores. This means that they are used less than 184 times for department B1VA and 163 times for B2VA. Table 2 gives an overview of the items that are used in less than one percent of the time and the amount of workload in minutes they represent.

B1VA				B2VA			
Code	Times used	Percen-tage	Work-load	Code	Times used	Percen-tage	Wo rk-
20	172	0,93	10	20	118	0,72	10
26	83	0,45	10	25	157	0,96	5
28	122	0,66	15	26	19	0,12	10
37	8	0,04	5	28	57	0,35	15
38	32	0,17	20	33	137	0,84	5
43	182	0,99	15	37	65	0,40	5
47	29	0,16	10	39	139	0,85	15
52	166	0,90	10	41	118	0,72	25
53	103	0,56	10	43	72	0,44	15
63	143	0,78	2	44	72	0,44	20
				46	26	0,16	30
				47	27	0,16	10
				52	29	0,18	10
				53	20	0,12	10
				54	19	0,12	15
				57	31	0,19	6
				58	52	0,32	20
				63	128	0,78	2

Table 2: The codes used in less than one percent of the scores.

Table 2 shows that many of the items almost which are never used are the same for both of the departments. In order to simplify the measurement instrument these codes should be deleted from the list of items used in the measurement instrument. The percentage in table 2 shows that all of these items are ticked in less then one percent of the time a nurse uses the measurement instrument. This means that when removing these items from the measurement instrument a small portion of the total workload measured gets lost compared to the actual workload, but removing these items makes the measurement instrument more straightforward and synoptic. In conclusion, items 20, 26, 28, 37, 43, 47, 52, 53 and 63 should be removed from the measurement instrument.

4.2.3 Items needed

In total three hundred twenty-three actions of nurses were timed and coded during the six days of observing. To investigate if items need to be added to the list of sixty four items currently in the measurement instrument the code 0 was added to the code list. During the observations every action performed by the nurse which could not be linked to one of the existing codes was coded with a zero. To get the insight needed in the content of these zero's a text field was added to the observation form in which a description of the action performed could be added. This means that thirty-eight percent of the actions performed by the nurse are not recorded in the measurement instrument. To get a better insight in the division of those codes the frequency of the observed codes is made visible in figure 6.

If there are items that possibly need to be added to the measurement instrument it is important to investigate what actions are currently coded with a zero. In table 3 the actions behind code zero are categorized and the number actions in the category are counted. The categorization is based on the three aspects of nursing work introduced in the conceptual model (figure 1) and divided into logical groups of activities.

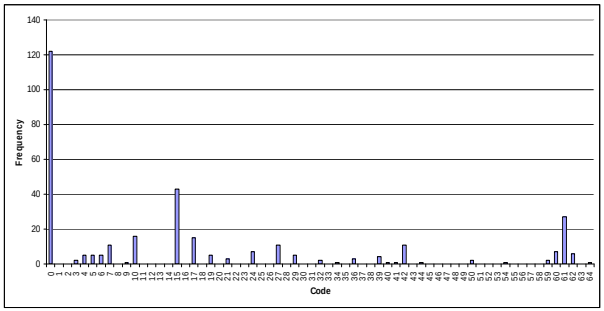

Figure 6: The frequency of codes observed during the research.

Category	Frequency
Patient care	20
Arranging something for a patient	9
Administrative tasks	33
Householding tasks	32
Cooperate with medical specialists	24
Cooperate with nurses	4
Cooperate with other staff	1
Total	123

Table 3: The content of code 0.

As mentioned in section 4.2.1 the measurement instrument focuses on direct nursing care activities. In table 3 it is shown that most of the zeros observed, hundred and three of the hundred twenty three, can be categorized under indirect patient care and non-patient related care activities. In order to make the current nursing workload measurement instrument valid these hundred and three activities need to be taken in to account when calculating the nursing workload over a period of time. This can be done by adding the actual items to the items to the measurement instrument, but most of the activities described cannot be linked to a particular patient. Therefore, a better choice would be to add a number of minutes to the calculation of the total nursing workload which represent indirect patient care activities and non-patient related care activities. As mentioned chapter 2.3 it is often assumed in other workload formulas that the total nursing workload consists for thirty-five percent of direct nursing care activities. If a choice is made to add a number of minutes to the calculation extensive research needs to be done to determine how many minutes a nurse spends for these category activities per time period.

Category 'Patient care'

The activities shown in table 4 together form the category 'Patient care' in table 3 and are activities that can be linked to a patient. These activities are the obvious candidates to be added to the measurement instrument. However, for all of these activities the questions should be asked if the activity occurs often enough to add as a new item and if it is possible to add the activity to another item by simply changing the description and the time. An activity can be added to an existing item when it is clearly linked to this item, because it is always performed after each other. For example 'prepare medication' can be added to the time spend serving medication, but we do not know if this is already calculated due to the lack of documentation.

Action	Frequency	New item	Added to other item
Help into bed	6		x
Getting acquainted with patient and family	2		
Prepare medication	4		x
Apply not mentioned medication	2		x
Help with daily hygiene	1		
Get extra food	1		
Move to other room	1		
Move to operation room	2		x
Prepare for operation	1		
Total	20	0	4

Table 4: The content of category 'Patient care'.

Looking at table 4 with these two questions in mind it can be concluded that none of the activities need to be added to the measurement instrument. Most of the activities can be covered by other items by a small change in description. Other activities in table 4 occurred only once during the observations and therefore it is probable that these items would be ticked more in less than one percent of the times a nurse fills in the measurement instrument. In conclusion, the following activities need to be added to existing items which can be found in figure 3:

- 'Help into bed' can be added to code 49,
- 'Prepare medication' can be added to 58, 59, 60 and 61,
- 'Apply not mentioned medication' can be added to 58, 59, 60 or 61,
- 'Move to operation room' can be added to code 4.

Other categories

Besides the activities in the category 'patient care' there are a few other activities behind code zero that are striking. In table 5 the most observed activities outside of the category 'Patient care' are shown. These activities are found in every other category besides the category 'Patient care' and are notable because of their high frequency. When looking at the conceptual model of nursing work in figure 1 it is clear that only one of the activities shown in table 5 can be categorized as direct patient care. The other activities, categorized in table 3 as 'administrative tasks' and 'cooperate with medical specialists', can be categorized as indirect patient care and non-patient care related based on the conceptual model.

Activity	Frequency
Change bedsheets	27
Assisting the doctor with daily round	20
Administration	12
Reporting	11
Scoring	6

Table 5: Activities besides category 'patient care'.

All of the activities mentioned in this table are performed daily. Therefore the choice needs to be made if these activities need to be new items or if the time needed for these activities needs to be taken into account in the calculation of the output of the measurement instrument. All of the activities, except 'Administration', are performed daily on or for every patient the nurse is responsible for. A sub-objective of this research is to make the measurement instrument simpler by grouping or removing some items in the list. Therefore, it is logical to research the exact time needed for these activities for each patient and add that number of minutes to the calculation of the nursing workload instead of adding new items that need to be ticked every time a nurse uses the measurement instrument.

4.2.4 Grouping of items

In order to try to make the measurement instrument simpler grouping of items currently used is a logical option. To determine which items can be grouped the correlation coefficients between all items in the historic data are calculated.

In table 6 the highest correlations between the current items of the measurement instrument are shown. In order to group some of these items together a high correlation coefficient is needed. A correlation coefficient of 0,90 or higher indicates that the two items closely vary together. Based on the information shown in table 6 it can be concluded that none of the items currently used can be grouped together. The items with a high correlation are logically connected to each other. For example it is obvious that a patient needing full help mobilizing also needs full help with washing. That the correlation is not high enough to propose grouping of these items is either a result that in practice the connection between these codes are not that obvious or that the nurses perform these items in one activity and therefore do not thick both items.

B1VA				
Code	Description	Code	Description	Correlation
48	Barriere / isolation nursing	5	Install patient	0,537
50	Full help with mobilizing	7	Full help with washing	0,598
39	Tube feeding	25	Nursing gastric hose	0,541
B2VA				
Code	Description	Code	Description	Correlation
27	Nursing Cad	7	Full help with washing	0,543
50	Full help with mobilizing	7	Full help with washing	0,514
54	Nursing tracheacanule	52	Sucking fauces	0,596
54	Nursing tracheacanule	53	Assist with bronchial toilet	0,512

Table 6: Correlation between current items

4.2.5. Conclusion

The most important conclusion of section 4.2 is that only direct patient care activities are used in this measurement instrument. This leads to an important question whether to add new items to the measurement instrument or add minutes, for indirect patient care activities and non-patient related activities, to the calculation of the nursing workload. Based on the results presented in section 4.2 it is concluded that adding minutes to the calculation of the nursing workload is the better option. This research focuses on determining to what extent content of the measurement instrument is valid, other research needs to be done to determine the exact amount of minutes that need to be added for each category.

4.3 The times

In this section the times used in the measurement instrument are evaluated. The data of the research cannot be used to evaluate every item of the measurement instrument. Some activities are observed several times and valid conclusions can be based on these observations. Other items are only observed a few times or never at all, conclusions cannot be based on these observations. Therefore, the research question regarding content validity focuses on the most observed items. These items can be used to compare the time in the measurement instrument and the actual times measured during the observations. Table 7 shows the ten most observed items. In appendix D the table with the complete comparison can be found.

The column 'Score' of table 7 shows the number of minutes every item represents in the current measurement instrument. The table shows that the averages are higher than the medians for most of the items, which means that the distribution of the measured times is positively skewed. Therefore, it is better to use the median for the comparison of the times used in the measurement instrument and the actual times. However, the average times can be used to calculate the average total workload over a period of time. When calculating the average total workload it is important to consider that an average is sensitive for extreme measures.

When comparing the score of an item to the median found in the observations it is clear that the current scores used in the measurement instrument do not represent the actual times spent by the nurses well. Furthermore, for most of the items the standard deviation is almost as high as the average time measured, meaning that sixty-eight percent of the times measured would lie between zero and twice the average time measured. This dispersion in the measured times leads to the conclusion that the measurement instrument in its current form cannot be used to get a good insight in the nursing workload over a short period of time.

This also means that based on this research no propositions can be made for a better matching time for an item.

Code	Score	Observations	Median	Average	Standard deviation
7	20	11	12,6	14,26061	9,894252
10	10	16	2,933333	3,626042	2,870987
15	7	43	2,95	3,129457	2,219891
17	3	15	0,883333	2,523333	3,012288
24	10	7	6,816667	11,83095	8,308398
27	5	11	2,083333	2,131818	0,900994
42	10	11	4	4,615152	3,384699
60	10	7	3,283333	3,57619	1,160147
61	10	27	2,066667	2,77963	3,242568
62	6	6	1,116667	0,986111	0,279453

Table 7: The ten most observed items.

However, it can be used to get an insight in the nursing workload over a longer period, because the deference between the actual average workload and the measured average workload will get smaller as an item occurs more in the dataset.

4.3.1 Conclusion

This section has clearly shown that the times used in the measurement instrument do not match the average times spent by the nurses for the ten most observed items. The dispersion of the measured times leads to the conclusion that it is hard to use the current measurement instrument for insight in the nursing workload over a short time-period. The measurement instrument can only be used to get a good insight in the average nursing workload over a longer period of time.

4.4 The users

In this section the behaviour of the users of the measurement instrument, the nurses, is investigated. Firstly, the usage of the measurement instrument is described based on the historic data of the measurement instrument. Secondly, the behaviour of the nurses is described based on observations during the use of the measurement instrument and the interviews done with every observed nurse.

4.4.1 The usage of the measurement instrument

In this subsection the usage of the measurement instrument in both wards is elaborated. Figure 7 and 8 show the usage per month of the measurement instrument for wards B1VA and B2VA. The first column of each month shows 'patient days', which is calculated as follows: the sum of the number of days every patient received care on the ward that month. The second column shows the number of entries in the measurement instrument.

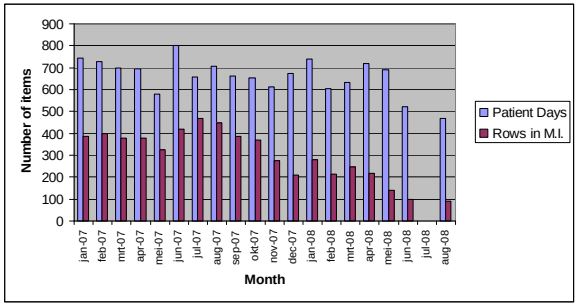

Figure 7: The number of patient scored per day at B1VA.

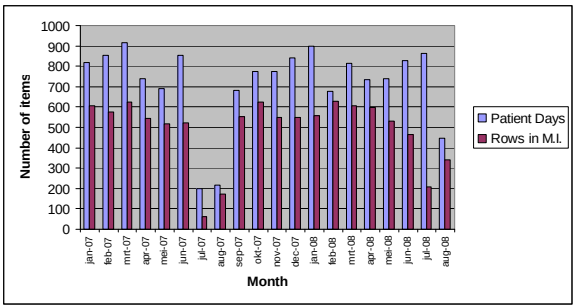

Figure 8: The number of patient scored per day at B2VA.

Both of these figures clearly show that the number of entries in the measurement instrument are structurally less than the number of patient days. If every patient was scored in the measurement instrument every day he or she was on the ward both of the columns should be exactly the same. The fact that this is not the case leads to the conclusion that not every nurse scores his or her patients in the measurement instrument every day.

4.4.2 The use of the measurement instrument

This section describes the most notable problems that occur when nurses use the measurement instrument. During the research several nurses using the measurement instrument were observed and at the end of every day of observing the nurse was interviewed. The data gathered with these two activities form the basis for this subsection. The most important finding in the interviews and observation is that the moment to use the measurement instrument is not part of the daily routine of the nurses. They quickly use the measurement instrument when they have a few minutes of spare time, which leads to a speedy and uninterested approach to scoring the patients in the measurement instrument. This is supported by the fact that every nurse mentions in the interview that they do not fully understand why it is important to use the measurement instrument.

Secondly, the choice which items to thick for differences in the use of the measurement instrument between the nurses. For some of the nurses it is unclear which item to choose and choices are therefore made based on individual interpretations of the description of an item. Nurses perform the same activities, but choose to score those activities in a different manner.

4.4.3 Conclusion

This section showed that the usage of the measurement instrument is structurally less than the actual patient days. Keeping the way of calculating the nursing workload in mind, this leads to an output in which the nursing workload is always portrayed much lower than it actually is. Furthermore, the interviews and observations have shown that the nurses have a lack of understanding of the measurement instrument and its objective.

5 Conclusions

In this chapter the research questions, as discussed in section 1.3.3, are answered. The questions focussed on the applicability, validity and reliability of the nursing workload measurement instrument. The same topics are used in this chapter describing the final conclusions based on the exploration of the current situation in chapter 3 and the results of the observations and interviews in chapter 4.

Applicability
In the current situation the lack of documentation prevents the use of the measurement instrument in any kind of decision making process. The measurement instrument can show different charts, but in order to use the measurement instrument in the decision making process the information needs to be understand and interpreted by a third party. This can only be done with good documentation of the development process and the choices made during this process.

Despite the fact that the applicability of the current measurement instrument is questioned, a few possibilities exist for the use of the current measurement instrument. At this moment the frequency of items ticked in the measurement instrument can be shown, indicating a possible change in the content of nursing on the thorax centre over a period of time. Furthermore, a trend in the nursing workload can be derived from the data collected with the measurement instrument. This trend shows whether the nursing workload is increasing or decreasing, despite the fact that times used for calculating the output are not accurate.

Validity
The validity of the measurement instrument is directed by two sub questions. These questions focus on the content validity and the construct validity of the measurement instrument. Two of the three requirements, mentioned in section 2.7, focus on the validity of the measurement instrument. Namely, the measurement instrument needs cover all the three aspects of nursing work and the operation times used in the measurement instrument need to match the actual operation times.

The content validity is investigated by the comparison between the time used in the measurement instrument for each item and the actual time measured during the observations. The times measured poorly match the 'scores' (used times) used when calculating the output of the measurement instrument. As result of the variability of measured times the measurement instrument is not usable for getting insight in the nursing workload of a short period of time. However, it can be used when looking at the nursing workload over a longer time period. The scores used in the measurement instrument still need improvement.

Construct validity focuses on the question whether the items on the measurement instrument cover the definition of nursing workload. At this moment the definition of nursing workload is not fully covered by the item list, since only direct patient care activities are listed. As mentioned in the results it is not necessary to add new items to the measurement instrument. The minutes spend on indirect patient care activities and non-patient care related activities need to be taken into account when calculating the nursing workload. The can be based on the types of patients that are scored. The same can be concluded for the direct patient care activities that were missing, these activities were performed on every patient every day. The minutes spend on the activities that are missing in the item list can be added to the calculation in order to keep the measurement instrument simple and understandable.

Reliability
The reliability of the measurement instrument is also covered by two sub questions, regarding the test re-test and inter-observer reliability. Reliability is one of the three

requirements of an applicable measurement instrument, described in section 2.7. To obtain useful output all of the nurses need to use the measurement instrument in the same manner

The test re-test reliability focuses on different times the same nurse uses the measurement instrument. In the results is shown that not every nurse uses the measurement as it supposed to be, the interviews and observations have shown a clear lack of understanding of the measurement instrument itself and the objective of the measurement instrument. This leads to a speedy and uninterested use of the measurement instrument, which consequently leads to an unreliable input in the measurement instrument.

The inter-observer reliability focuses on the difference in the use of the measurement instrument between the nurses. The results have shown that the usage of the measurement instrument is structural less than the actual patient days what would imply an output in which the nursing workload is always portrayed much lower than it actually is.

In conclusion, the applicability of the current measurement instrument is questionable. All of the requirements for an applicable measurement instrument are not met. The research has shown that the list of items do not cover the three aspects of nursing work, the times used for every item are questionable and the nurses do not use the measurement instrument in the same manner. However, when the measurement instrument is improved the management will have a great tool to support their budget requests and get insight in the nursing work on their wards.

6 Suggestions for improvements

In this chapter improvements are described in order to make the current measurement instrument applicable. Improvements are needed on all of the three requirements for an applicable measurement instrument in order to develop a measurement instrument that can be used in the decision making process.

Firstly, in order to improve the validity of the current measurement instrument two things need to be done. Further research needs to be started to determine the time spend on every item in the measurement instrument and activities that are not covered with the item list need to implemented in the calculation of the total nursing workload. The exact time spend per time period on these activities also needs to be researched.

Secondly, in order to improve the reliability of the measurement instrument it is important to get everyone to use the measurement instrument after the day shift and to get everyone to use the measurement instrument in the same manner. A large part of this can be done by teaching the nurses how to use the improved measurement instrument and by showing them what is done with the output, making them more enthusiastic to use the measurement instrument. However, a hundred percent using rate will almost never be accomplished. Therefore, the input of every day needs to be supplemented with the workload of an average patient for every patient on the ward that did not receive a score. This needs to be done every day, because the bed occupation data of days in the past are not available in the current measurement instrument.

Finally, in order to improve the usage of the measurement instrument, besides the usage by nurses, it is important to change the output to simple charts in which the total nursing workload is shown over a time period. When the use of this output is fully integrated in the decision making process other charts and options can be added to the output, but it is important to keep the output of the measurement instrument simple and understandable in order to make it easy to show the collected information to third parties.

In conclusion, to improve the measurement instrument to make it useful and take advantage of its possibilities, the following steps need to be taken:
Extensive research needs to be done in order to obtain valid scores to use in the measurement instrument. This is important for the items in the list as well as the activities added in the calculation.
Documentation needs to be written in which the design of the measurement instrument can be described and motivated.
After implementing these improvements the nursing staff needs educated in the use of the measurement instrument in order to get reliable results in the future.

Bibliography

Adcock R. & Collier D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. American Political Science Review 95(3).

Albie-Gibbons A. (1986) Mental health acuity system: The measure of nursing practice. Journal of Psychosocial Nursing and Mental Health Services 24, 16- 20.

Arthur T. & James N. (1994). Determing nurse staffing levels: a critical review of the literature. Journal of Advanced Nursing 19, 558-565.

Audit Commission (1991). The virtue of patients: Making best use of ward nursing resources. HMSO, London.

Arthur T. & James N. (1994). Determining nurse staffing levels: a critical review of the literature. Journal of Advanced Nursing 19, 558–565.

American Nurses Association Policy Statement (2003). Retrieved from: www.nursingworld.org/. In Prescott P.R., Socken K.L., Castor A.H, Thompson R.O. & Phillips C.Y. (1991) The Patient Intensity for Nurses Index: a validity assessment. Research in Nursing & Health 14, 213–221.

Benefield L.E. (1996) Productivity in home healthcare: assessing nurse effectiveness and efficiency: part one. Home Healthcare Nurse 14 (9), 699–706.

Brady A.M., Byrne G., Horan P., Griffiths C., MacGregor C. & Begley C. (2007). Measuring the workload of community nurses in Ireland: a review of workload measurement systems. Journal of Nursing Management 15, 481–489.

Bridel J. (1993). Why measure workload? Professional Nurse 8(6), 362-365.

Cambell, T., Taylor, S., Callaghan, S. & Shuldham, C. (1997). Casemix type as a predictor of nursing workload. Journal of Nursing Management 5, 237–245.

Caplan R.D. & Jones K.W. (1975). Effects of workload, role ambiguity, and type A personality on anxiety, depression, and heart rate. Journal of Applied Psychology 60, 713–719.

Carr-Hill R.A. & Jenkins-Clarke (1995). Measurement systems in principle and in practice: the example of nursing workload. Journal of Advanced Nursing 22, 221-225.

Coles J. & Jenkins L. (1992). Computerized nursing workload systems. Senior Nurse 12(4), 5-7.

Cronbach L.J. (1990). Essentials of Psychological Testing. HarperCollins, New York.

Davidhizar, R., Mallow, G. & Bechtel, G. (1998). A patient classification system for the chronic psychiatric patient. Australian and New Zealand Journal of Mental Health Nursing 7, 126-133.

Department of Health & Social Security (1986). Health Services Management. Resource Management (Management Budgeting) in Health Authorities. HMSO, London.

Edwardson S.R. & Giovannetti P.B. (1994). Nursing workload measurement systems. Annual Review of Nursing Research 12, 95-123.

Endacott R. & Chellel A. (1996). Nursing dependency scoring: measuring the total workload. Nursing Standard 10(37) 39– 42.

Freeman S., Shelley G. & Gay M. (1999). Measuring services: a district nursing dependency tool. Nursing Standard 13(47), 39–41.

Gibbs, I., McGaughan, M. & Griffiths, M. (1991). Skill mix in nursing: a selective review of the literature. Journal of Advanced Nursing 16, 242–249.

Gliner J. A., Morgan G. A., & Harmon R.J. (2001). Measurement Reliability. Journal of the American Academy of Child & Adolescent Psychiatry 40(4).

Grandfield S. (1992). Do more with less: strategies for improving productivity in community nursing. Health Care Supervisor 11(1) 37–42.

Henderson V. (1966). The Nature of Nursing, A Definition and its Implications for Practice, Research and Education. Macmillan, New York.

Hernandez C. & O'Brien-Pallas L. (1996). Validity and reliability of nursing workload measurement systems: review of validity and reliability theory. Canadian Journal of Nursing administration 9(3), 32–50.

Hughes M. (1999). Nursing workload: an unquantifiable entity. Journal of Nursing Management 7, 317–322.

Hurst, K. (2003). Estimating the Size and Mix of Nursing Teams. Nuffield Institute for Health, Leeds.

L.ohr K.N., Aaronson N.K., Alonso J., Burnam A.M., Patrick D.L., Pen-in E.B. & Roberts J.S. (1996). Evaluating Quality-of-Life and Health Status Instruments: Development of Scientific Review Criteria. Clinical therapeutics 18(5).

Malloch, K. & Conovaloff, A. (1999). Patient classification systems, Part 1: the third generation. Journal of Nursing Administration 29, 49-56.

McKenna H.P., (1995). Nursing skillmix substitutions and the quality of care: an exploration of assumptions from the research literature. Journal of Advanced Nursing 21, 452–459.

Morgan G. A., Gliner J. A. & Harmon R.J. (2001). Measurement Validity. Journal of the American Academy of Child & Adolescent Psychiatry 40(6).

Morris R., MacNeela P., Scott A., Treacy P. & Hyde A. (2007). Reconsidering the conceptualization of nursing workload: literature review. Journal of Advanced Nursing 57(5), 463–471.

National Health Service Management Executive (1989). Nursing Information Requirements: Identification and Computerisation. NHS Management Executive, Leeds.

Needham J. (1997). Accuracy in workload measurement: a fact or fallacy? Journal of Nursing Management 5, 83–87.

Needleman, J., Buerhaus, P. & Mattke, S. (2002). Nurse staffing levels and quality of care in hospitals. New England Journal of Medicine 346 (22), 1715–1722.

O'Brien A.J., Abas M., Christensen J., Nicholls T., Hekau A. & Vanderpyl J. (2002). Nursing Workload Measurement in Acute Mental Health Inpatient Units. Health Research Council of New Zealand, Auckland.

O'Brien, G., (1986). The intuitive method of patient dependency. Nursing Times, Occasional Paper 82 (8), 57–61.

O'Brien-Pallas L., Irvine D. & Peereboom E. (1997). Measuring nursing workload: understanding the variability. Nursing Economics 15, 171–182.

O'Brien-Pallas, L., Thomson, D., Alksnis, C., & Bruce, S. (2001). The economic impact of nurse staffing decisions: Time to turn down another road? Hospital Quarterly 4(3), 42-50.

Polit D. & Hungler P. (2003). Nursing Research: Principles and Methods. Lippincott, Philadelphia.

Prescott P.R., Soeken K.L., Castorr A.H., Thompson K.O. & Phillips C.Y. (1991). The Patient Intensity for Nursing Index: a validity assessment. Research in Nursing and Health 14, 213–221.

Seago J.A. (2002). A comparison of two patient classification instruments in an acute care hospital. Journal for Nursing. Administration 32, 243–249.

Sovie M. & Smith T. (1986). Pricing the nursing product: charging for nursing care. Nursing Economics 4, 216–226.

Stillwell J.A. & Hawley C. (1993). Nursing dependency and the costs of nursing care. Journal of Nursing Management 1, 113-117.

Walker A.R. & Whynes D.K. (1990). The costing of nursing care: a study of 65 colorectal cancer patients. Journal of Advanced Nursing 15, 1305–1309.

Whitney, J.A. & Killien, M.G., (1987). Establishing predictive validity of a patient classification system. Nursing Management 18, 80–86.

Wright-Warren P. (1986). 'Mix and Match' a Review of Nursing Skill Mix. NHS Management Board, London.

Appendix A Observation scheme

Code	Omschrijving
0	niet beschreven
Opname/Ontslag	
1	Anamnese Lang
2	Anamnese Kort
3	Ontslag
4	Patiënt terughalen van IC/REC./CCU
ADL	
5	Patiënt installeren / Materiaal aanbieden
6	Gedeeltelijk hulp: rug-voeten e.d.
7	Volledige Hulp
8	Ontharen (pre-OK)
9	Steunkous
10	Hulp bij po of toiletbezoek
11	Incontinent
12	Laxeren
Monitoring	
13	> daags controle (pols,temp,rr)
14	3 en 4 x daags controle
15	1 en 2 x daags controle
16	Bloedafname
17	Wegen
18	ECG
19	Bloedsuiker dagcurve
20	Pacemaker drempelen (assisteren
21	Saturatie meten
22	Telemetrie stickers verschonen
23	Telemetrie met bewaking

Code	Omschrijving
Mentaal	
35	Patiënt is verward
36	Psychosociale ondersteuning fam. + pat.
37	PTA scorelijst
38	Pat. Voorlichting hartfalen/hartinfarct
Voeding	
39	Sondevoeding geven
40	Enige hulp bij eten en drinken
41	Volledige hulp
Wondzorg	
42	Standaard
43	Spoelsysteem, verzorgen van
44	Spoelen en Furacine gazen
45	Frequente verbandwisseling i.v.m. lekkage
46	VAC-wissel
47	Zwachtelen
48	Barrière / isolatie verpleging
Mobiliseren	
49	Met enige hulp
50	Met volledige hulp
Ademhaling	
51	Hulp bij ophoesten
52	Keelholte uitzuigen
53	Bronchiaal toilet, assisteren bij
54	Tracheacanule, verzorgen van

Belijning

24 Thoraxdrains, verzorgen van
25 Maagslang, verzorgen van
26 Maagslang ingebracht
27 CAD, verzorgen van
28 CAD ingebracht
29 Perifeer infuus, verzorgen van
30 Perifeer infuus ingebracht
31 Centrale veneuze catheter, verzorgen van
32 Redonse drain, verzorgen van
33 Epiduraal, verzorgen van
34 Pacemaker draden, verzorgen van

Decubites

55 Duo matras, verzorgen van pat. op
56 Primo matras, verzorgen van pat. op
57 Blokkenmatras, verzorgen van pat. op

Medicatie

58 >2 x dd IV medicatie
59 1 en 2 x dd IV medicatie
60 >2 x dd orale e.a. medicatie
61 1 en 2 x dd orale e.a. medicatie
62 Sprayen
63 Eigen beheer (alle medicatie)
64 Zuurstof toediening, verzorgen van

Appendix B Code scheme

Code	Tijd	Bed	Overig		Code	Tijd	Bed	Overig

Appendix C Interview scheme

Introduction
Today I observed you during a complete dayshift. As mentioned before the shift started I did this for my research regarding the nursing workload measurement instrument. This research focuses on whether or not the nursing workload measurement instrument in its current form is useful for the management of the thorax centre. Besides the observations I want to ask every nurse I observed a few questions in order to get insight in how the staff feels about the nursing workload measurement instrument.

Questions

What is your opinion of nursing workload measurement instrument in its current form? Open question

Do you think the nursing workload measurement instrument in its current form is usefull for this ward? Open question

How often do you use the nursing workload measurement instrument?
Never / Most of the time not / Most of the time / Always

When you use the nursing workload measurement instrument are there items in the list that you never use?
Item number ….., why?

Are there items in the nursing workload measurement instrument of which the description needs to be changed to make them more clear to you?
Item number ….., why?

Do you miss any items in the current nursing workload measurement instrument that need to be added to align the instrument in a better way with the actual nursing work?
Open question

At the end of this research project the goal is to have a good working nursing workload measurement instrument. Do you think this nursing workload measurement instrument has any added value for this ward? Open question

Closure
This was the final question of the interview. I want to thank you for letting me observe you today and participating in the interview.

Appendix D Instrument and observation data

Code	Item	Score	Observations	Median	Average	Standard deviation
1	anamnese lang	30	0			
2	anamnesr kort	15	0			
3	ontslag	10	2	3,125	3,125	0,991667
4	pat. terughalen v. IC/CCU	20	5	12,03333	12,45	6,486173
5	pat. installeren	10	5	2,65	3,793333	2,341097
6	ged.hulp, rug-voeten ed	10	5	4,95	7,903333	7,274448
7	volledige hulp	20	11	12,6	14,26061	9,894252
8	ontharen	10	0			
9	steunkous	5	1	2,466667	2,466667	0
10	hulp bij po	10	16	2,933333	3,626042	2,870987
11	incontinent	15	0			
12	laxeren	5	0			
13	meer dan 4x dd contr	20	0			
14	3 en 4 x dd contr	12	0			
15	1 en 2 x dd contr	7	43	2,95	3,129457	2,219891
16	Bloedafname	7	0			
17	wegen	3	15	0,883333	2,523333	3,012288
18	ECG	6	0			
19	Bloedsuiker dagcurve	10	5	1,783333	2,03	0,647851
20	Pacemaker dremp	10	0			
21	Saturatie meten	4	3	1,533333	1,283333	0,47629
22	telemetrie stickers versch	3	0			
23	telemetrie met bewaking	15	0			
24	Thoraxdrains	10	7	6,816667	11,83095	8,308398
25	Maagslang, verzorgen	5	0			
26	Maagslang ingebracht	10	0			

Code	Item	Score	Observations	Median	Average	Standard deviation
27	CAD, verzorgen van	5	11	2,083333	2,131818	0,900994
28	CAD, ingebracht	15	0			
29	Perifeer infuus, verzorgen	5	5	2,316667	4,593333	4,282621
30	Perifeer infuus, ingebracht	10	0			
31	Centr.veneuze cath., verz.	8,5	0			
32	Redonse drain, verzorgen	5	2	4,133333	4,133333	2,633333
33	Epiduraal, verzorgen van	5	0			
34	pacemaker draden verz	5	1	0,916667	0,916667	0
35	Patient is verward	60	0			
36	Psychosoc.onderst. fam	30	3	2,633333	3,211111	2,268844
37	PTA score lijst	5	0			
38	pat. voorlichting hartf/inf	20	0			
39	Sondevoeding geven	15	4	13,10833	14,05417	5,529697
40	Enige hulp bij eten en drinken	15	1	4,6	4,6	0
41	Volledige hulp	25	1	21	21	0
42	Wondverz. standaard	10	11	4	4,615152	3,384699
43	Spoelsysteem, verzorgen van	15	0			
44	Spoelen en Furacine gazen	20	1	5,983333	5,983333	0
45	Freq. verbandwisseling ivm lek	25	0			
46	VAC-wissel	30	0			
47	zwachtelen	10	0			
48	barriere /isolatie verpl	30	0			
49	Mob. met enige hulp	15	0			
50	Mob. met volledige hulp	20	2	3,666667	3,666667	0
51	Hulp bij ophoesten	10	0			
52	Keelholte uitzuigen	10	0			
53	Bronchiaal toilet, ass. bij	10	0			
54	Tracheacanule, verz. van	15	1	13,5	13,5	0
55	Duo matras, verz. v. pat. op	10	0			
56	Primo matras, verz. v. pat. op	10	0			

Code	Item	Score	Observations	Median	Average	Standard deviation
57	Blokkenmatras, verz. v. pat. p	6	0			
58	> 2x dd IV medicatie	20	0			
59	1 en 2 x dd IV medicatie	10	2	7,566667	7,566667	2,816667
60	> 2x dd orale medicatie	10	7	3,283333	3,57619	1,160147
61	1 en 2 x dd orale medicatie	10	27	2,066667	2,77963	3,242568
62	spayen	6	6	1,116667	0,986111	0,279453
63	Eigen beheer (alle medicatie)	2	0			
64	Zuurstof toediening, verz.van	5	1	1	1	0