CROSS-CULTURAL ADAPTATION AND VALIDATION OF A SCORE FOR EVALUATING THE QUALITY OF INPATIENT CLINICAL NOTES

By

Sonia E. Benitez

A MASTER‟S THESIS

Presented to the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of
Master of Science

December 2013

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's Thesis of

Sonia Elizabeth Benitez

"Cross-cultural adaptation and validation of a score for evaluating quality of inpatient clinical notes"

Has been approved

[redacted]

Thesis Advisor: Paul Gorman, MD

[redacted]

Member: Michael Chiang, MD

[redacted]

Member: Justin Fletcher, PhD

[redacted]

Member: Paula Otero, MD

[redacted]

Member: Diego Giunta, MD

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

TABLE OF APPENDICES

LIST OF ABBREVIATIONS

CEPI: Comité de Etica de Protocolos de Investigación

DMICE: Department of Medical Informatics & Clinical Epidemiology

ECG: Electrocardiogram

EHR: Electronic Health Record

EHRs: Electronic Health Records

FOCUS: Friends of Overseas Citizens and University Students

GRRAS: Guidelines for Reporting Reliability and Agreement Studies

GS: Gold Standard

HIBA: Hospital Italiano de Buenos

ICC: Intra-class Correlation Coefficient

ICU: Internal Care Unit

IRB: Institutional Review Board

JCAHO: Joint Commission on Accreditation of Healthcare Organizations

MMSE: Mini-Mental State Examination

NYP Hospital: New York–Presbyterian Hospital

OHSU: Oregon Health & Science University

PDF: Portable Document Format

PDQI-9: Physician Documentation Quality Instrument.

POS: Palliative care Outcome Scale

Acknowledgments

Preparing and completing my master‟s thesis has been the most gratifying and incomparable learning experience of my life, and it would not have been possible without the continued support of many people. I wish to thank my thesis advisor, Dr. Paul Gorman, from Oregon Health & Science University (OHSU), for his advice and assistance in keeping my progress on schedule. He has given me constant encouragement and guidance throughout my time as his student. My grateful thanks also to Dr. Diego Giunta, my tutor at Hospital Italiano de Buenos (HIBA) in Argentina, for his valuable and constructive recommendations during the planning and development of this project. I would also like to express my gratitude to the rest of the attendings and staffs who work with him, especially to Dr. Lourdes Posadas Martinez, for her help during the planning of this research work. For their kindnes and support, I would like to thank Dr. Paula Otero, Dr. Justin Fletcher and Dr. Michael Chiang, members of my thesis committee. For providing funding for this project through the Informatics Training for Global Health grant I would like to acknowledge the National Institutes of Health‟s Fogarty International Center.

The valuable support of members of HIBA helped me complete this project. I would like to express my gratitude to Dr. Daniel Luna and Dr. Fernán Quirós, for seeing all that I am capable of and helping me move ahead in my career. I would also like to express my deep gratefulness to Nancy Orrego and Mariela Lage –both part of the team of Dr. Santiago Wassermann, coordinator of the Terminology Area– for their assistance with data collection and the logistics of the project. My thanks also to the rest of the members of the Health Informatics Department. For supporting me along the process, my appreciation is extended to

Amparo Dominguez and her team in the Billing Department. I would like to express my deep gratitude to all the Internal Medicine physicians for their contribution to this research: they became directly involved in the project, generously giving their time to evaluate the score, or performing the evaluations. In addition, they indirectly participated by encouraging me to keep on with the project. My grateful thank you also to the secretaries of the Internal Medicine Department.

This achievement was also possible with the support of students, faculties and staffs of Department of Medical Informatics & Clinical Epidemiology of OHSU (DMICE) and the amazing people of Portland, Oregon. I wish to thank my fellow students whom I met during my experience at DMICE, principally Leeann Stahn, Akiwumi Olobumi, and Tracy Edinger. Thank you to the faculties, particularly Holly Jimison, Jayashree Kalpathy-Cramer, Joan Ash, Michelle Hribar, Byung Park, and the program director, Dr. William Hersh, and to the staffs, especially Diane Doctor, Alicia Nam, Elizabeth Brentnall Williams, and Lynne Schwabe. Thank you for your patience, guidance, advice, support, and recommendations. I would also like to acknowledge the support of Sally and Tom Downing, Anabel Lopez Salinas, Cherryl Dillon and her family, all the members of "Friends of Overseas Citizens and University Students" (FOCUS), and Maria Elsa Gray and her family. My life in Portland would not have been the same without all of you. The friendship and laughter we have shared made me feel at home.

Finally, thanks to my mom, Teresa Leguizamon, for believing in me and cheering me on. Without the strength of your support, I would not have had the courage to do it.

Abstract

Introduction: Communication between health care providers is an important factor for the continuity of patient care. In medical records, clinical notes (admission notes, progress notes and discharge summaries) become important for this purpose. The inappropriate use of features of electronic health records (EHRs), such as copy & paste, produces "unreadable" documents with redundant information that reduces the quality of clinical notes. Within this framework, it is essential to assess the quality of clinical notes with a view to implementing mechanisms to improve them.

This study proposes to develop a new version of the Physician Documentation Quality Instrument (PDQI-9). This scale evaluates the quality of clinical notes for the purpose of physician communication. Because of a measuring instrument should be reliable and valid beyond the original research population, a cross-cultural validation of the score is critical.

Objective: To do a cross-cultural adaptation and validation of the Spanish version of the PDQI-9 score.

Methods: This project analyzed the use of the PDQI-9 in Spanish, at Hospital Italiano de Buenos Aires, in electronic clinical notes. The process of cross-cultural validation involved two phases. The first phase comprised the translation and cross-cultural adaptation of the instrument with the participation of an expert panel and a group of Internal Medicine physicians for the pre-test evaluation.

The second phase, included the assessment of internal consistency, intra-rater reliability, inter-rater reliability and the evaluation of the criterion validity of the adapted

instrument. The evaluation of an expert committee represented the gold standard (GS). **Results:** The translation and adaptation of the score were completed successfully. For the pre-test evaluation, the average time of evaluation of the clinical notes was 17.27 (SD 9.53, min 5, max 46) minutes.

In the validation process, 26 physicians were raters and 8 physicians were GS. They performed 898 evaluations, and the evaluation was done in 10.28 (SD 6.81, min 1, max 70) minutes in average. The overall Cronbach's Alpha coefficient was 0.92. For the inter-rater reliability evaluation, the global intra-class correlation coefficient (ICC) for 898 evaluations was 0.92 (95% CI 0.91 - 0.93, p<0,001). The test retest was performed in 30/898 (3.3%) of all the clinical notes evaluated. In terms of criterion validity, the correlation coefficient was 0.85 (p <0.001).

Discussion: The objective of a cross-cultural study is to obtain equivalence. Equivalence was achieved through the cross-cultural adaptation (semantic, idiomatic, experiential, conceptual, and content) and the validation (criterion) process. Psychometric measurements of the adapted instrument were similar to those of the original instrument. Due to cultural differences, the recommendation is to do an extra cross-cultural study in other Spanish speaking countries. Additional research may answer whether the attributes in the PDQI-9 score are sufficient to evaluate the quality of clinical notes in other specialties or if those attributes fit other purposes of the EHR.

Conclusion: The cross-culturally adapted Spanish version of the PDQI-9 instrument is reliable and valid for Internal Medicine physicians, to be used in Spanish-speaking countries, especially in Argentina, at Hospital Italiano de Buenos Aires.

Introduction

Communication between health care providers is an important factor for the continuity of patient care. In the inpatient setting, communication failures can lead to patient management, omission, diagnosis, treatment, and commission errors. Those incidents might be due to poor transmission and exchange of information and interpersonal relationship problems. The transfer of information between professionals may be through verbal orders, handoffs or medical records. Within medical records, narrative components such as clinical notes (admission notes, progress notes and discharge summaries), become important for this purpose (1-2). Admission notes and discharge summaries allow the reconciliation of information between levels of care. The admission note contains a summary of the past medical history and is a snapshot of the reason for hospitalization. Progress notes allow daily communication and transition between different providers during hospitalization (i.e., a discharge summary is not used when a transfer from the internal care unit (ICU) to a non-ICU setting happens; in this case, a progress note can replace it). Finally, the discharge summary sums up the relevant information from hospitalization.

The medical record is a means of communication among clinicians, but this can be an ineffective way to communicate, if the notes in the medical records are not clear (or missing) and no other communication can be established. This leads to an instance of disagreement in the healthcare team (3) and the dynamic characteristics of the medical care promote the discontinuity of care. In this scenario, organizations such as the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) have paid particular attention to the discontinuity of patient care as a negative factor for patient safety especially in the transfer

of information (4). According to the JCAHO, communication and information management problems were the most commonly identified root causes of sentinel events, especially in delay in treatment events, leading to death or permanent loss of function from 2004 to 2012 (5).

Besides communication, the information contained in the medical record allows research, education of new providers, and supports the care provided as an administrative, billing and legal document (6) in addition to contributing to the provision of proper patient care (Figure 1: Reasons for using medical records).

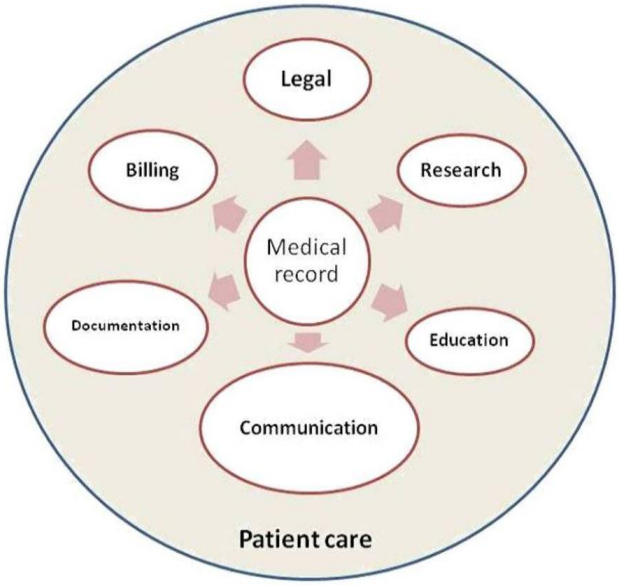


Figure 1: Reasons for using medical records

Unlike paper-based records, the electronic health record (EHR) has the advantage of supporting extended activities (See Figure 2: Dynamic functions of the EHR). The EHR allows quality management and outcome reporting, supports healthcare-related activities –

including clinical decision support– and ultimately “automates and streamlines the clinician's workflow” (7).

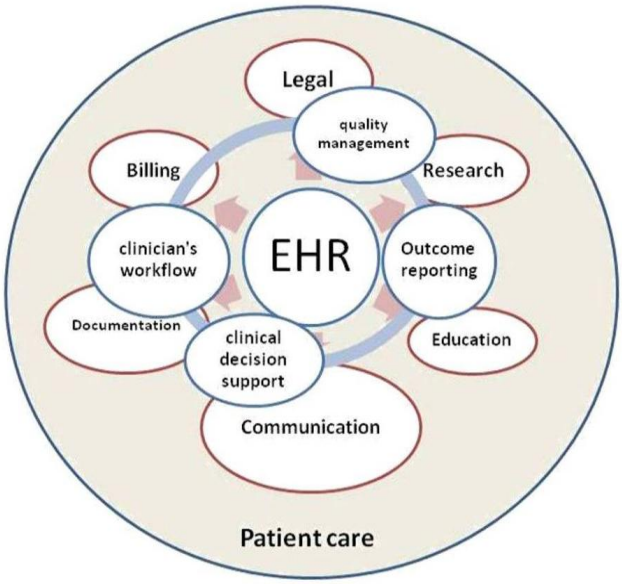


Figure 2: Dynamic functions of the EHR

The use of the EHR has benefits such as ensuring accessibility and availability of information, visualization of data in different ways, provision of contextual information, communication with other professionals and patients, among others, and has solved several handwriting-related issues such as legibility (8). However, the EHR has generated new concerns. The introduction of the EHR, especially clinical narrative documents, in the daily workflow may have altered direct communication between physicians (1). The inappropriate use of functionalities such as copy & paste or the use of templates for creating clinical notes impact on readability, producing "unreadable" documents, outdated and lengthy progress notes with redundant information, all of which reduces their *quality* (1, 9-10) and affects the exchange of the information therein contained (2). Lengthy progress notes are associated

with “information overload” (11-12) as well as “information chaos” (13) and can impact on patient care through the propagation of inaccurate information (9, 14).

Therefore, within this framework, it is essential to assess the quality of clinical notes with a view to implementing mechanisms designed to improve them. However, “quality” is a subjective concept and depends on users´ needs. In addition, studies related to the evaluation of quality of electronic clinical notes have been scarce, and tended to be oriented to multiple and heterogeneous stakeholders without a clear purpose of the clinical note established or level of patient care. Unlike that research work, Stetson et al. (15-16) have been working on a score called Physician Documentation Quality Instrument (PDQI-9) that evaluates the quality of inpatient clinical notes with the clear purpose of enhancing communication among physicians.

There are increasing EHR implementation initiatives in Latin America and these initiatives have particularities in terms of language and culture. In addition, when a score is used in a different country –a language and culture other than their original ones– it is necessary to conduct a cross-cultural adaptation (17) to use that score locally (18-19). The cross-cultural process implies adapting and checking whether the instrument maintains the ability to measure the same characteristics in accordance with the original design. This is essential to compare results (20). Therefore, to use the PDQI-9 score in our setting, it is necessary to develop a local language (Spanish) version equivalent to the original one, and measure the local validity (21). The purpose of this study is to do a cross-cultural adaptation of the PDQI-9 score and the subsequent validation at Hospital Italiano de Buenos Aires (HIBA), in Argentina.

General Objectives

1. To do a cross-cultural adaptation and develop a Spanish version of the PDQI-9 score.

2. To do a validation of the adapted version of the PDQI-9 score.

Specific Objectives

1. To develop an instrument through the creation of a Spanish version of the pre-tested PDQI-9 score.

2. To perform the reliability assessment.

3. To do the validity evaluation.

Background

Quality is a vague and subjective concept. Juran (22) defined *high-quality data* to be data "fit for use in their intended operational, decision-making, planning, and strategic roles." For English (23) “quality” is “fitness for all purposes made of the data, including the likely future uses”. English also defined “quality information” as “consistently meeting knowledge worker and end-customer expectations.” In other words, quality information is the expectation “to have the necessary resources available to carry out the work” (22). According to the American Society of Quality, quality is “a subjective term for which each person or sector has its own definition”. In the manufacturing arena, a product or service has quality when it is free of deficiencies, but for the purpose of this study the definition of

quality is represented by the characteristics of a product (or service) that satisfy users‟ needs (9, 24).

The study of the quality of the medical record is not new and there have been many attempts to evaluate it. Most of the times, attributes of quality are completeness or how accurate the information is (25-26). Nevertheless, this view would be associated with the evaluation of the information in a “structured way”. The quality of the narrative component of the medical record would have some other attributes that can describe them better, beyond “accuracy” or “completeness”. The following paragraphs contain an overview of the research done about the evaluation of the quality of electronic clinical notes.

Quality of Clinical Notes

Research into the evaluation of quality of electronic clinical notes is still limited. At the time of this research, there are four groups working on the quality of clinical notes: Hanson, Hammond, Shen and Stetson groups. Table 1 presents a comparison of studies about the quality of clinical notes.

Table 1: Comparison of studies about quality of clinical notes

	Stetson *(PDQI-9 score)*	**Hammond**	**Shen**	**Hanson**
User perspective	Physicians	Multiple; neither researchers nor patients	Hammond	Multiple; no researchers
Purpose	Communication	-	-	-
Level of care	Inpatient	Inpatient/ Outpatient	Inpatient/ Outpatient	Outpatient
Score	Yes	Yes	Yes	No
Attributes	Up-to-date Accurate Thorough Useful Organized Comprehensible Succinct Synthesized Internally consistent	Informativeness Easy to read Understandability Trust in the information	Templates Headings Inserted objects	Conciseness Sufficiency of information Self-explanatory Relevant Clear Readable Current Accurate Organized

Hanson et al. (27) evaluated clinical notes in the ambulatory setting, following the perspective of multiple users: clinicians, nurses and ancillary staff, administrators; neither patients nor researchers were included. Using the grounded theory approach, a qualitative study was conducted. Three principal topics defined the quality, the content and the systems supporting the quality of clinical notes. A quality clinical note should be written as if a story were being told. It should have conciseness and sufficiency of information, be self-explanatory, relevant, clear, readable, current, accurate and organized. The aim of this research group is to create a score in the future; and the purpose of the clinical note (communication or other) is not clear.

Hammond and collaborators (28) studied the perceptions of quality by conducting semi-structured interviews to physicians, nurses, physician assistants and administrative

personnel who use the Veterans Affairs EHR system. The researchers created a ten-attribute instrument and evaluated twelve documents from a single de-identified patient health record. The researchers used a simulated EHR instead of printed copies. Physicians, nurse practitioners and physician assistants were more stringent when they evaluated some dimensions of quality in the clinical notes. Those dimensions were informativeness, reading ease, understandability, and trust in the information. The purpose of the clinical note was not specified either. Neither completeness nor accuracy was evaluated quantitatively. This study reinforced the notion that quality depends on the role or task that the health care provider performs.

In the 2010 i2b2/VA challenge, Shen et al. (29) used the instrument developed by Hammond and his team of researchers, this time focusing on the assessment of readability and informativeness of specific aspects of the clinical notes. Those aspects are structural such as templates, headings and inserted objects in the clinical notes. Four reviewers evaluated 246 clinical notes (progress notes and discharge summaries). In general, the structural aspects had a positive association with document quality. For instance, the use of headings makes it easier to find information; the use of templates, however, may confuse the reader, depending on the design and organization of the template.

Unlike the studies mentioned, Stetson et al. (15) focused on physicians and the use of the clinical notes as a means of communication among physicians. The authors identified a list of 22 attributes concerning the quality of clinical notes of inpatients for the purpose of assessing physician communication, reviewing the literature and interviewing clinical experts. In a second study, after supplementary experts‟ review and performing factors analysis, they reduced the amount of attributes. The resulting score contains 9 items, to

which an item of overall impression was added. The evaluation of the PDQI-9 score was with Internal Medicine physicians at the New York–Presbyterian (NYP) Hospital, a New York City-based academic institution. Two groups of physicians evaluated the new instrument in admission notes, progress notes and discharge summaries. In one group, the expert physicians represented the gold standard; in the other group were the attending physicians and residents. The reliability and validity evaluation consisted in assessing criterion-related validity, discriminant validity, internal consistency and inter-rater reliability. This score does not evaluate the structure, the components or the sections of the clinical note, because the intention of the authors was to allow the use of the score at outpatient level as well (16).

There is an apparent benefit in using the PDQI-9 instrument. However, the original instrument is in the English language. The following paragraphs describe the advantages and disadvantages of adapting a score or developing one "de novo".

Rationale and Purpose of the Study

In Latin America there are many measuring instruments that have been developed in another language and then adapted to Spanish, especially in the health care domain. For example, the Mini-Mental State Examination (MMSE) is a questionnaire in English to screen cognitive impairment originally presented by Folstein et al. (30) and now considered the gold standard. This questionnaire was translated into Spanish by Lobo et al. (31). Some experts in psychological tests prefer the adaptation when the test is valid, as determined by research, and when the adaptation follows a strict process of evaluation, allowing further comparative studies. Others prefer the development of new measuring instruments rather than their adaptation/translation because: 1) the adaptation of the instrument can lead to

errors, 2) a translation that overlooks cultural differences biases results; 3) the phenomenon under study may not be well represented by the test (32). Creating a score from scratch would have the advantage of abstracting the quality attributes that would be locally valid. However, adapting and validating a validated score makes it possible to have a validated construct and save time.

In this master project, the construct is the set of attributes of the original instrument. A priori, the construct of quality created by Stetson‟s team would be appropriate for the new setting or targeted culture, because there are similarities between the NYP hospital and HIBA. HIBA is a university hospital located in the city of Buenos Aires. The Internal Medicine area has a head of service, attending physicians, chief resident and residents in the organizational structure. Residents are supervised by attending physicians, make morning rounds, use handoffs, use HIBA‟s EHR, and are on duty, among other tasks. They have a weekly journal club and the medical literature reviewed is, essentially, in English. Nevertheless, and according to Guillemin et al. (17), every time an instrument is intended to be used in another country, language and culture, it is necessary to conduct a cross-cultural adaptation and not just a simple translation. The following table (Table 2: Different scenarios for conducting a cross-cultural study, adapted from (17)), summarizes the scenarios, when it is necessary to just translate, or to translate and adapt.

Table 2: Different scenarios for conducting a cross-cultural study, adapted from (17)

Change in			Adaptation	
Culture	***Language***	***Country of use***	***Translation***	***Cultural adaptation***
no	**no**	**no**	**no**	**no**
yes	**no**	**no**	no	yes
yes	**no**	yes	no	yes
yes	yes	no	yes	yes
yes	**yes**	**yes**	**yes**	**yes**

A measuring instrument should be reliable and valid, for the original and the new research population. The validity of the instrument in a new environment is tied to a process of "cultural adaptation" (21). Cross-cultural validation implies adapting and checking whether the instrument maintains the ability to measure the same features or characteristics according to the original design and is essential for comparing results (20). Following the advantages mentioned about score adaptation, and because the instrument will be used in another country, language and culture, the intention of this study is to do a cross-cultural adaptation of the instrument to the Spanish language.

Materials and Methods

The following sections deal with the research questions, term definitions, ethical considerations, framework, design, setting, population, and the process of cross-cultural adaptation and validation.

Research Questions

This project analyzed the use of the PDQI-9 instrument in a different language (Spanish), in an institution of different size (HIBA), using a different EHR (Italica). Therefore research questions were:

1. Is the Spanish version of the PDQI-9 score valid and reliable in an institution of a different size, with a different EHR?

2. Is the Spanish version of the PDQI-9 score valid and reliable for internal medicine physicians?

Among the sub problems detected were:

a. ***To translate*** the PDQI-9 score to its Spanish version.

b. ***To cross-culturally adapt*** the PDQI-9 score to its Spanish version.

c. ***To validate*** the PDQI-9 score in its Spanish version.

The physicians at HIBA were assumed to know how to use the EHR and how to create clinical notes.

Term Definitions

- Clinical notes: These are the narrative components of the EHR, represented by admission notes, progress notes and discharge summaries.

- Communication between physicians: The process by which patient information is exchanged among physicians through verbal orders, handoffs, medical records or other systems (adapted from (1, 32)).
- Quality: The characteristics of a product (or service) that satisfy users‟ needs.
- Cross-cultural adaptation: This process takes into account the cultural context, language and differences in the perception of the phenomenon under study in the populations where the instrument applied (33).

Ethical Considerations

This cross-cultural validation study did not involve any additional risk to the patient or physicians. Physicians used the adapted score with de-identified clinical notes. The study was conducted in full accordance with national and international laws: the Helsinki Declaration of the World Medical Association and the Good Clinical Practice Rules (E6 ICH). All study data was treated with utmost confidentiality and anonymously, with access restricted to authorized personnel only in connection to the study, in accordance with the National Law on Personal Data Protection 25,326 (Habeas Data Act) (34).

The Institutional Review Board (IRB) of Hospital Italiano (Comité de Etica de Protocolos de Investigación, CEPI (35)) approved protocol number 2027 on March 18, 2013. An English version of the approval was sent to the IRB of OHSU. The IRB of OHSU approved the protocol (Short Study Title: "Cross cultural adaptation of the PDQI-9 score", IRB Number: IRB00009564, Review Category: Exempt) on May 10, 2013. Between April 15 and May 10, the master‟s candidate did pre-recruiting tasks, including presentations of the project.

Framework

This project followed the taxonomy created by Gremy & Degoulet (36). The dimensions of the technology assessment proposed are:

1. The Medical Informatics Applications: In this dimension, the clinical notes as a part of the EHR (Clinical System Domain).

2. The taxonomy of people or network of actors represented by physicians.

3. The technology assessment itself: at the technical level, the domain will be the quality of patients‟ records.

4. Time: This dimension was not under evaluation in this study.

Design

The original PDQI-9 instrument (See Appendix 1: The PDQI-9 Instrument) was cross-culturally adapted and validated to develop a Spanish version. In 2012 Dr. Peter D. Stetson, original researcher and principal investigator in the development of the PDQI-9 score, authorized by e-mail the use of this score.

Research Setting

This study was conducted at HIBA, a non-profit tertiary health care academic center with a capacity of 750 beds, 500 home care patients, and 24 outpatient care centers. In 1998, HIBA first introduced a Healthcare Information System (HIS). The EHR is web-based, problem-oriented and patient-centered. In addition, HIBA is working on improving the processes that involve accreditation by the Joint Commission International.

Research Population

This study included electronic clinical notes (admission notes, progress notes and discharge summaries) and physicians from the Internal Medicine Department following the original research protocol. The participation of physicians was free and voluntary. The Internal Medicine Department will receive a mobile computer as an incentive for participating in this project and concurrent projects.

Process of Cross-cultural Adaptation and Validation

The process of cross-cultural adaptation and validation involved two phases: the first one comprised the translation and cross-cultural adaptation of the instrument. The outcome of this phase was a Spanish version of the score and involved an expert committee and a group of internal medicine physicians for the pre-test evaluation. The second phase comprised the evaluation of the reliability and validity of the Spanish version. This phase included the assessment of internal consistency, intra-rater reliability, inter-rater reliability and the evaluation of the criterion validity of the adapted instrument. The validity of the adapted instrument was measured by comparing the average total score using the adapted score with the average overall impression of the GS group. An expert committee represented the GS. For every phase, the clinical notes (admission notes, progress notes and discharge summaries) were obtained by random and stratified selection.

Physicians used the Spanish version of the PDQI-9 instrument with clinical notes (admission, progress notes and summary discharge). The original instrument (see Appendix 1: The PDQI-9 Instrument) has nine items evaluating attributes of quality using a Likert scale

(1 to 5: “1” means the attribute is not present in the clinical note; “5” means the attribute is present in its highest expression in the clinical note) with a total score range of 9-45. The higher is the score, the better is the quality. The following diagram, Figure 3: Translation, cross-cultural adaptation and validation process (adapted from (33)), depicts the research workflow:

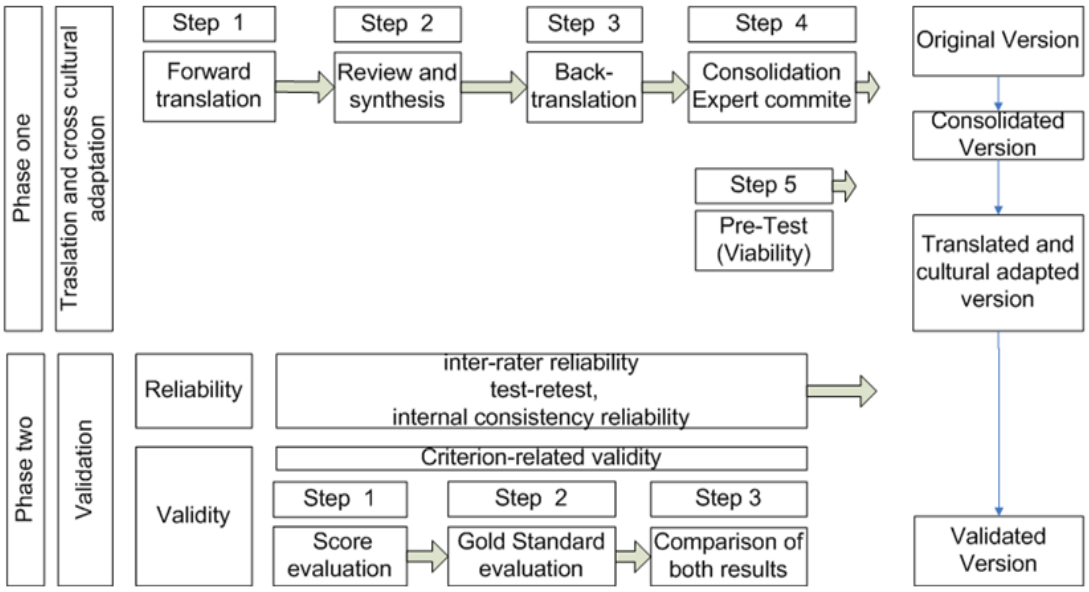


Figure 3: Translation, cross-cultural adaptation and validation process (adapted from (33))

The following paragraphs describe the selection of translators, back translators, experts and the clinical notes.

Selection of Translators

Two key representative physicians of the Internal Medicine Department and two members of the thesis committee of the master‟s candidate identified six people (three

physicians and three laypersons) qualified to translate the instrument: fluent in the target language with good understanding of the original language (18), i.e. Spanish native speakers , with good understanding of English. Eligible translators received the invitation to participate by e-mail; three of them agreed to participate and, finally, two of them translated the instrument.

Selection of Back-translators

The master‟s candidate asked two members of her thesis committee, three DMICE-OHSU PhD candidates, three DMICE-OHSU MBI candidates , two DMICE-OHSU faculties, and one Family Medicine physician from OHSU, to identify people who qualified to back translate (to translate it from Spanish to English) the instrument. The back translators should be fluent in the original language, with good understanding of the target language (18); that is, native speakers of the English language, with good understanding of Spanish. In addition, the master‟s candidate reviewed the OHSU hospital website to characterize eligible people. The Family Medicine physician suggested contacting the "National Hispanic Medical Association" (http://www.nhmamd.org/). This process resulted in the identification of sixteen people who were skilled enough to act as back translators (thirteen physicians, one Medical Informaticist, one medical student and one Master in Public Health). Eligible back translators were e-mailed invitations to take part in the project, and five of them agreed to participate. Two of them did not reply additional e-mails, one person back translated the instrument partially, and finally two of them back translated the instrument prior to the experts‟ meeting.

Selection of the Expert Committee

For expert physicians to be deemed qualified for adapting the instrument, selecting or reviewing the clinical notes, they had to 1) have experience in resident training and 2) be involved in the evaluation of residents in the areas of communication skills and use of EHRs. They might be current or past training directors, key internal medicine representatives or from subspecialties of internal medicine. These physicians were required to have completed the Internal Medicine residency, and been active at least 3 years after completing their residency.

Instrument Adaptation Experts

The expert committee entrusted with adapting the instrument included four internal medicine physicians (adapters). The master‟s candidate, one member of the thesis committee of the master‟s candidate, one chief resident in Internal Medicine and two key representative physicians of the Internal Medicine Department identified ten attending physicians eligible to be members of the instrument adaptation expert committee. They received the invitation to join it by e-mail.

Experts for the Gold Standard Group

The expert committee in charge of reviewing clinical notes without the use of the instrument included at least seven internal medicine physicians, following the original research. The master‟s candidate, one member of the thesis committee of the master‟s candidate, one of the chief residents in Internal Medicine and two key representative

physicians of the Internal Medicine Department identified thirteen attending physicians eligible as GS. They received the invitation by e-mail.

Clinical Notes Selection Expert

A key representative physician of Internal Medicine was invited to participate in the selection of clinical notes (selectors). This physician participated neither in the adaptation nor in the validation phase. One member of the thesis committee of the master‟s candidate suggested his participation. The eligible physician accepted to participate.

Corpus of Clinical Notes

This study used a set of de-identified copies of clinical notes from selected EHRs. A key representative physician of Internal Medicine independently reviewed the EHRs and selected the clinical notes. He reviewed the clinical notes following his own criteria in accordance with the objective of the score (evaluation of quality according to the communication of physicians for continuity of care). He received a list of the EHRs to be reviewed.

Inclusion Criteria for Clinical Notes

The list of all patients *hospitalized at the Internal Medicine* Department in HIBA, *without any referral*, *with at least three days of hospitalization*, from February to March 2012 and June 2012 was requested from the Biostatistics Area of HIBA. The reason for selecting those dates was that the inpatient EHR system was under implementation (with e-prescribing, nurse chart, electronic admission notes electronic and progress notes already implemented)

between 2010 and May 2011. The documents available before that date are scanned documents. On the other hand, the selection of documents from 2013 could affect the evaluation because the participants might recognize the patients. In addition, the list of all patients *hospitalized at the Internal Medicine* area in HIBA, *without any referral*, *with at least three days of hospitalization* from June 2012 were selected for looking for clinical notes could probably have lower quality due to the recent arrival of the new medical residents. The figure below (Figure 4: Selection of clinical notes flowchart) shows the clinical note selection process.

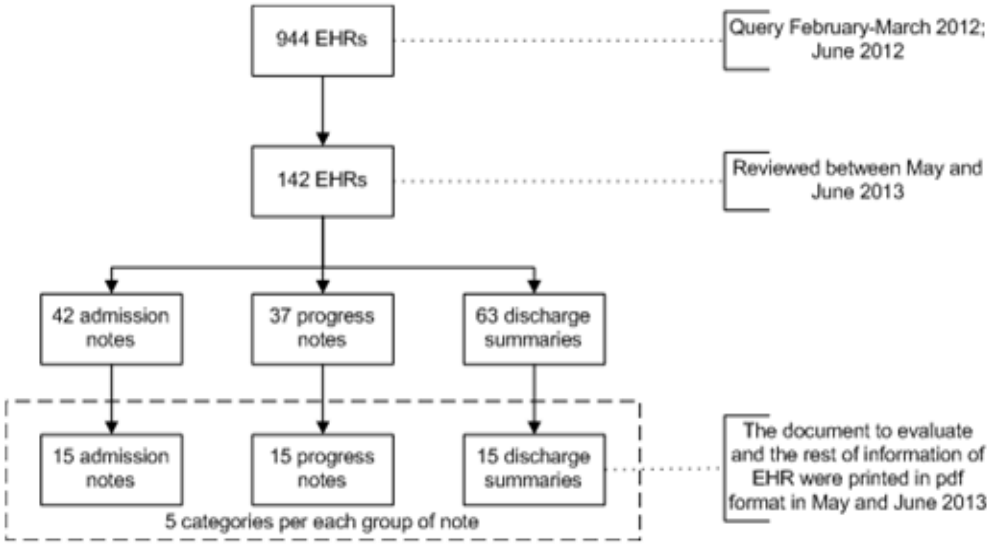


Figure 4: Selection of clinical notes flowchart

Every patient on the list was randomly numbered from zero to one, using the random function of Excel ® .The random numbers were sorted from low to high before the review process. The random number generator from www.random.org set the progress note to

evaluate. The website of Random.org recommends the "True Random Number Generator" (see Figure 5: True Random Number Generator).

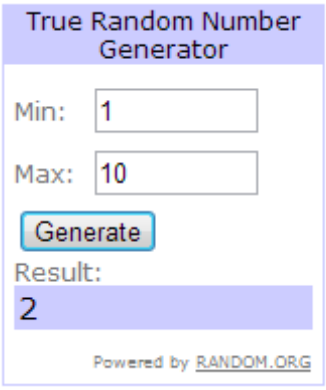


Figure 5: True Random Number Generator

The minimum (Min.) number was 1 (one) and the maximum (Max.) number was the total hospitalization days. The number in the result section was the progress note to be evaluated. In the example, the progress note under evaluation was issued on day two of hospitalization. The selector physician received the date of the clinical note to evaluate. Based on the list and accessing the EHR, the selector physician had to grade clinical notes using a scale from one to ten, and five categories were defined from one to five. Category number one contained clinical notes rated one to two; category number two comprised clinical notes rated three to four; category number three contained clinical notes rated five to six; category number four included clinical notes rated seven to eight; and category number five contained clinical notes rated nine to ten. The objective was to have at least three admission notes, three progress notes and three discharge summaries (all documents from different EHR) in each category in order to have a spectrum of the clinical notes.

The clinical notes and the medical record were printed in Portable Document Format (PDF). PDFcreator® was the software used to this end. The documents were de-identified using the Infix PDF Editor ® demo version. The resulting de-identified medical records contained sections in chronological order including the admission note, progress notes, and the discharge summaries as well as previous discharge summaries or any other information from other levels of care (e.g., ambulatory), if necessary. The de-identified medical record also contained laboratory data and other relevant data such as Electrocardiogram (ECG) interpretations or image reports. These documents were in the same order as they were within the EHR. For every phase, clinical notes were obtained by random stratified selection. Five admission notes, five progress notes and five discharge summaries were selected for the adaptation phase, and ten admission notes, ten progress notes and ten discharge summaries were selected for the validation phase.

Sample Size

Physicians entrusted with evaluating the clinical notes using the adapted instrument were residents or attendings from the Internal Medicine specialty or subspecialties thereof. Seven Internal Medicine physicians participated in the adaptation phase: four physicians were members of the experts committee, and three physicians were part of the pre-test sub-phase. The number of physicians required for the validation phase was established following the original protocol: seven Internal Medicine expert physicians acted as the GS (not participating in the adaptation phase) and 24 internal medicine physicians (16).

1. Phase 1: Translation and Cross-cultural Adaptation

The outcome of this phase was a Spanish version of the score, which was obtained through the following steps:

1.1. Step 1: Forward Translation

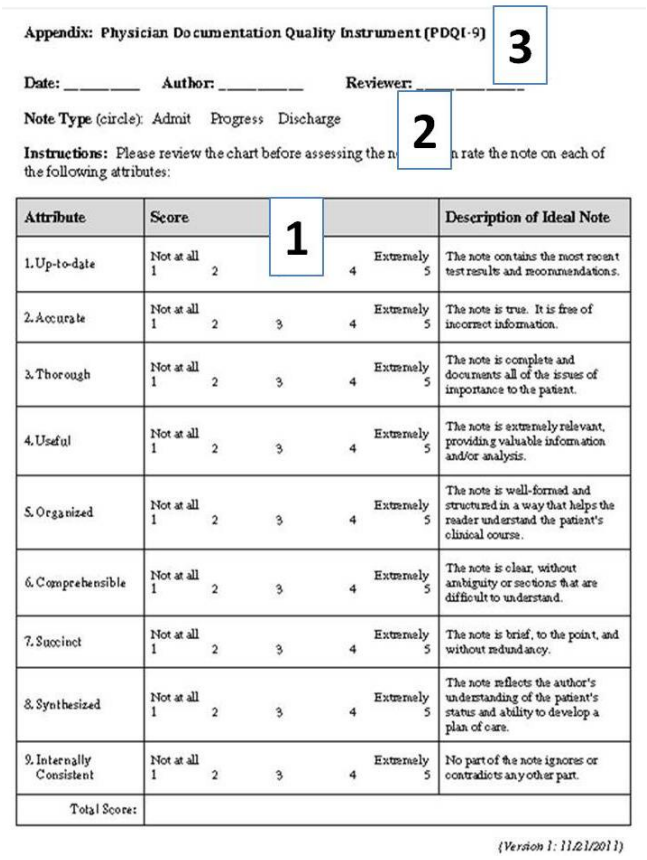
Two native Spanish speakers from Argentina who were fluent in English translated the original version of the score from English to Spanish. One of the translators was an internal medicine physician, who had a proficiency level of English, and the other translator was a bilingual executive secretary.

1.2. Step 2: Review and Synthesis

The objective of this step was to solve issues related to **inter-comprehensibility**, **ambiguity**, **understandability** and the **restriction on the range of the response**. **Inter-comprehensibility** means that one term has the same meaning to all the physicians; by **ambiguity** physicians understand that the term can have more than one meaning; **understandability** of every item as well the accompanying explanatory texts means that they have to be easy to understand. Finally, **restriction on the range of the response** means that raters may not answer an item because they do not understand the item or the item is not applicable to the evaluation. Experts proposed appropriate improvements to the format, if necessary.

The expert committee, including the master‟s candidate, reviewed the two translated versions and synthesized them into one, the beta version. All of the members of

the expert committee had Spanish as native language; they were from Argentina and had a good level of knowledge of English. The PDQI-9 instrument was divided into three parts (see Figure 6): part number 1 included the attributes, the score and the description of the ideal note; part number 2 contained the header (date, author, reviewer, type of note and instructions); part number 3 included the title.



Appendix: Physician Documentation Quality Instrument (PDQI-9) 3

Date: _____ **Author:** _____ **Reviewer:** _____

Note Type (circle): Admit Progress Discharge 2

Instructions: Please review the chart before assessing the n n rate the note on each of the following attributes:

Attribute	Score					Description of Ideal Note
1. Up-to-date	Not at all 1	2	3	4	Extremely 5	The note contains the most recent test results and recommendations.
2. Accurate	Not at all 1	2	3	4	Extremely 5	The note is true. It is free of incorrect information.
3. Thorough	Not at all 1	2	3	4	Extremely 5	The note is complete and documents all of the issues of importance to the patient.
4. Useful	Not at all 1	2	3	4	Extremely 5	The note is extremely relevant, providing valuable information and/or analysis.
5. Organized	Not at all 1	2	3	4	Extremely 5	The note is well-formed and structured in a way that helps the reader understand the patient's clinical course.
6. Comprehensible	Not at all 1	2	3	4	Extremely 5	The note is clear, without ambiguity or sections that are difficult to understand.
7. Succinct	Not at all 1	2	3	4	Extremely 5	The note is brief, to the point, and without redundancy.
8. Synthesized	Not at all 1	2	3	4	Extremely 5	The note reflects the author's understanding of the patient's status and ability to develop a plan of care.
9. Internally Consistent	Not at all 1	2	3	4	Extremely 5	No part of the note ignores or contradicts any other part.
Total Score:						

(Version 1: 11/21/2011)

1

Figure 6: The PDQI-9 instrument divided in three parts for a better evaluation: 1) attributes, score and description of ideal note; 2) header (date, author, reviewer, type of note, and instructions); 3) title

1.3. Step 3: Back translation

In this step, two different native English speakers who were fluent in Spanish independently back translated the beta version from Spanish to English. The back translators were e-mailed the beta version. One of the back translators was an Internal Medicine physician and the other one was a Master of Public Health. Both back translators had English as their native language and had studied Spanish. The Master of Public Health had been an exchange student in Buenos Aires and studied at the University of Buenos Aires a long time ago. To prepare the final version of the back translated document, she asked a clinician for help. The Internal Medicine physician had studied Spanish; however, he became fluent by practicing when he did his medical residency at Santa Clara Valley Medical Center, in San Jose, California, United States. The back translators did not know about the objectives or details of the project and were not part of the research either.

1.4. Step 4: Experts Consolidation

The objective of this step was to reformulate and consolidate a new version. There was consensus that the aim was to achieve semantic, idiomatic, and conceptual equivalence. The expert committee compared the original version against the back translated versions and the original version, the beta version, and the consolidated version. This process addressed issues concerning: 1) instrument format and writing style, 2) proper use of synonyms; 3) restriction on the range of the response; 4) need for training; 5) easiness to get the final score; 6) the time of application.

For the second meeting, experts were e-mailed a back translated version of the beta version before the meeting. The other back translated version was given to them at the meeting. Tasks were to review the back translations, make any suggestions geared towards improving attribute understanding, the description of the ideal note, the instructions and the title, and create the consolidated version. During the meeting, the experts and the master‟s candidate reviewed all of those aspects. The master‟s candidate collected the suggested changes and coordinated the two- hour meeting.

The same expert committee reviewed the two back translated versions and compared the back translated versions with the beta version. The expert committee also contrasted the original instrument and the back translated versions, they highlighted and discussed inconsistencies related to back translated words. The final outcome was the consolidated version, created according to the committee´s suggestions.

1.5. Step 5: Pre-test

In this step, participants evaluated the clinical notes and use of the consolidated version of the instrument. The master‟s candidate conducted semi-structured interviews on any aspect that had proved to be difficult to understand. This phase was intended to evaluate the quality of the translation, the cultural adaptation, the applicability of the score and the time needed to evaluate the clinical notes. The purpose was to achieve experiential equivalence. Readability of the descriptions of the ideal notes in the consolidated version was measured using readability metrics such as the Fernández Huerta scale (Spanish adaptation of the Reading Ease Score) (37), the Flesch-Szigriszt Index (35) and the PMOSE/IKIRSCH document readability formula (38). The Fernández

Huerta and Flesch-Szigriszt indices were obtained using the INFLESZ software (39), and the PMOSE/IKIRSCH was calculated using the guide developed by the Health Literacy Study groups of the Harvard School of Public Health (40).

2. Phase 2: Validation

The objective was to evaluate whether the translated and adapted instrument was valid and reliable. The validation phase comprised the following steps: 1) evaluation of clinical notes using the Spanish version of the PDQI-9 score; 2) evaluation of clinical notes using the GS; and 3) comparison of the results obtained using the Spanish version of the PDQI-9 score and the GS evaluation of the clinical notes.

2.1. Reliability Evaluation

2.1.1. Internal Consistency

The objective was to assess whether the items in the Spanish version of the PDQI-9 score have a good degree of correlation, for every type of clinical notes, using the Cronbach's Alpha coefficient.

Intra-rater Reliability (Test-Retest)

The aim was to evaluate whether the measurement was stable (if a similar result was obtained) when the instrument was applied a second time to the same document. The time defined to conduct the second evaluation was at least seven days after the first evaluation. Ten percent of the clinical notes were re-assessed to evaluate

test-retest reliability with the same physician: one admission note, one progress note, and one discharge summary.

2.1.2. Inter-rater Reliability

The objective was to evaluate the reproducibility of the Spanish version of the PDQI-9 score among different physicians, using the intra-class correlation coefficient for each score among the reviewers.

2.2. Validity Evaluation

This step was designed to assess the criterion validity of the adapted instrument.

2.2.1. Criterion Validity

The objective of this step was to evaluate whether the adapted instrument was valid by calculating the Pearson correlation coefficient between the average overall impression score of every clinical note and the average total PDQI-9 score of the same note. Results obtained (using the adapted score) were compared with the overall impression as determined by the experts (GS). The "overall impression" was an overall evaluation of the quality using a range of one to ten (1-10), according to a common grading scale in Argentina (41). Higher score means better quality.

Data Collection

Because physicians worked in the same facility, and to avoid the passing of information from one to another, seven groups were created. The thirty selected and de-identified clinical notes and the associated medical records were randomly classified with a letter: a, b, c, ch, d, e, f, g, h, i, j, k, l, ll, m, n, ñ, o, p, q, r, s, t, u, w, y, x, z, aa. After that, the de-identified clinical notes and the medical records were randomly numbered from one to thirty. The distribution of the numbers of clinical notes and medical records was different in every group. Participants did not know the original letter of the medical record and the same medical record could have a different number depending on the evaluation group.

Participants were e-mailed a link to access thirty folders containing the de-identified clinical notes (admission note, designated progress notes and discharge summary) and the associated medical records numbered from one to thirty, using the Dropbox® cloud storage service technology. Participants were asked to review the de-identified medical records as if they were "seeing the patient" for the first time, according to the original research. Physicians evaluated the quality by using the adapted instrument, while the GS evaluated the same clinical notes using the overall impression item. Upon request, raters received printed copies of the form with the adapted score. The GS received a form without the adapted score. Raters were free to use the hospital or other internet networks to conduct the evaluation. The time of evaluation was initially set in two weeks. Participants were asked to complete two evaluations at different time points: the second evaluation was to be one week after the first one.

Score Administration

Physicians were e-mailed the shared folders (Dropbox®) containing the clinical notes and medical records (pdf versions) and customized forms to fill in. Physicians got customized forms containing the initials of the reviewer, the type of clinical note and the number of medical record, in addition to the score and instructions. They had to complete the day and time of the review, and rate the clinical note. The rest of the information was pre- filled. Customizations were done to reduce fill-out times and to minimize the chances of reviewers to incur errors when evaluating clinical notes. Upon their request, physicians received a hard copy of forms with the score.

Statistical Analysis

Internal consistency of the adapted instrument was estimated using the Cronbach‟s Alpha (α) coefficient ranging from 0 to 1. The test-retest reliability was estimated using the Intra-class Correlation Coefficient (ICC). Criterion validity was measured using the Pearson coefficient of correlation (r). Statistical analyses were performed using the SPSS Statistical Software (SPSS 17.0 for Windows) and STATA. Table 3 summarizes the number of participating physicians, clinical notes and medical records as well as the proposed analysis according to the stages.

Table 3: Summary of the methodology

Phase	Sub-phase	Statistical analysis	Physicians	Clinical Notes			Medical Records
				Admission	Progress Notes	Discharge Summary	
Adaptation	Translation		1	-	-	-	-
	Review/ Consolidation		4	-	-	-	-
	Pre test		3	5	5	5	15
Validation	Reliability	Inter-observer reliability; Test-retest reliability: Intra-class Correlation Coefficient	24	10	10	10	30
		Internal consistency: Cronbach"s					
	Validity	Criterion validity: Pearson coefficient (r)	7 (24)				
Total			39	15	15	15	45

Results

This structure of the adaptation phase reports is based on "The Palliative care Outcome Scale (POS) Manual for cross-cultural adaptation and psychometric validation" (42). Reliability and validity process reporting followed the "Guidelines for Reporting Reliability and Agreement Studies" (GRRAS) (43), and the "International Commission Guidelines for test translation and adaptation" (44). The implementation of this study was from May to August of 2013 (See Figure 7: Timeline of the cross-cultural adaptation and validation of the Spanish version of the PDQI-9 score), based on the following objectives:

1. Cross-cultural adaptation: creation of the Spanish version of the PDQI-9 score.
2. Validation of the Spanish version of the PDQI-9 score: reliability and validity evaluation.

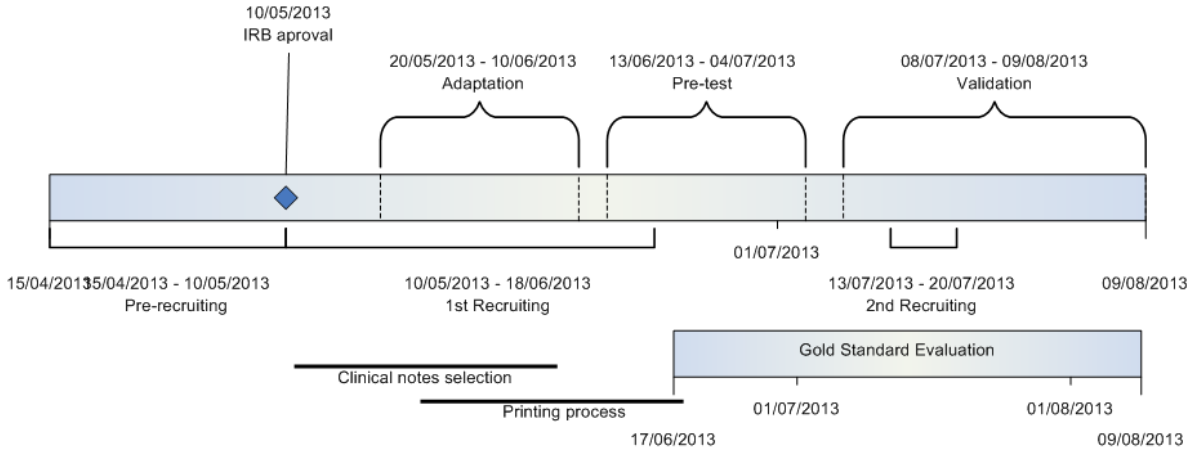


Figure 7: Timeline of the cross-cultural adaptation and validation of the Spanish version of the PDQI-9 score

The following paragraphs describe the results according to the objectives mentioned above.

1. Phase 1: Cross-cultural adaptation and creation of the Spanish version of the PDQI-9 score

The adaptation (phase one) of this research was carried out between May and June 2013, its outcome being the Spanish version of the score. The following diagram (Figure 8) depicts the phase workflow, and the sections below describe the process:

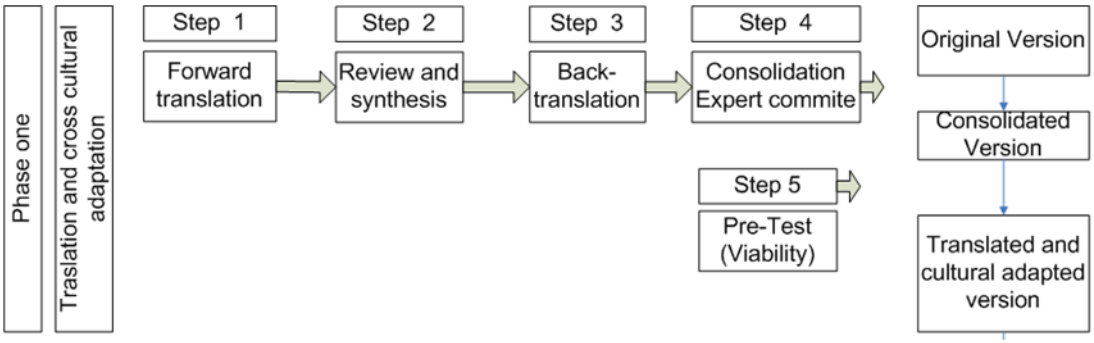


Figure 8: Phase one: Translation, cross-cultural adaptation process (adapted from (33))

1.1. Step 1: Forward Translation

The outcome of this first step comprised two independent Spanish versions of the PDQI-9 instrument. (See Appendix 4: Forward Translation of Attributes and Description of Ideal Notes, and Creation of the Beta Version; and see Appendix 5: Header and Title Forward Translation and Creation of the Beta Version.)

1.2. Step 2: Review and Synthesis

The PDQI-9 instrument comprised the following sections: section number 1 includes the attributes, the score and the description of the ideal note; section number 2 includes the header (date, author, reviewer, type of note, and instructions); section number 3 includes the title (See Figure 6 in Materials and Methods section). All items required discussion, and consensus by experts solved the inconsistencies. The translation of the word "note", from the concept of "clinical note", was "nota". Since this word, however, is not common in the Spanish medical lexicon, the experts chose the word "registro" instead. They agreed that "registro" conveyed the meaning of the word "note" in Spanish. Another translation could be "documento"; however, according to the experts, this word would be more associated with a legal record or the documentation process and not necessarily with the communication process. Experts preferred "registro" to "registro médico" (medical record) because the medical context should provide enough clarification.

Experts agreed that the attributes should be in the **adjective** form, wherever practicable, and highlighted and discussed every discrepancy among translated words. The addition of a word to the beta version occurred with the agreement of experts. The original version of the attributes are: "up-to-date", "accurate", "thorough", "useful", "organized", "comprehensible", "succinct, "synthesized", and "internally consistent". Finally, the attributes of the first version (beta version) in Spanish were: "actualizado" (up-to-date), "preciso" (accurate), "completo" (thorough), "útil" (useful), "organizado" (organized), "comprensible" (comprehensible), "conciso" (succinct), "sintetizado"

(synthesized), and "internamente coherente" (internally consistent). The ones around which there was more debate were "preciso" (accurate), "completo" (thorough) and "sintetizado" (synthesized).

Experts reviewed every description of the ideal notes, and rearranged words and sentences when necessary. The translations of the title were confusing due to the ways in which words had been arranged. At the suggestion of the master"s candidate and after discussing the matter, experts agreed to add the word "Evaluación" (evaluation) so as to improve title understandability. Finally, the title was defined as "Instrumento de Evaluación de la Calidad de la Documentación Médica". There was not a significant discussion about the Likert scale. Experts and translators were e-mailed the beta version and they agreed on the beta version. (See Appendix 3: Evaluation of Forward Translations and Creation of the Beta Version Report.)

Review Process

During the initial review process, experts did four reviews. They compared the original version with the forward translated version 1 and version 2; both forward translated versions, and the original version with the beta version. A matrix of agreement among versions was constructed. (See Appendix 9: Comparison of the Original English Version with the Backward Translation, and Appendix 11: Matrix of Agreement Among the Different Versions of the PDQI-9.)

1.3. Step 3: Back-translation

The outcome of this step was to have two independent back translated versions of the beta version. See Appendix 7: Backward Translations of Attributes and Description of Ideal notes, and Creation of the Consolidated Version, and Appendix 8: Header and Title Backward Translations and Creation of the Consolidated Version.

1.4. Step 4: Expert Consolidation

At this step, and after having reviewed the back translated versions, experts agreed with most back translated words. The attributes of the consolidated Spanish version were: “actualizado” (up-to-date), “preciso” (accurate), “completo” (thorough), “útil” (useful), “organizado” (organized), “comprensible” (comprehensible), “sucinto” (succinct), “interpretado” (synthesized), and “internamente coherente” (internally consistent).

Experts decided that “sucinto” (succinct) conveyed the concept described in the ideal note. On the other hand, the word “sintetizado” (synthesized) did not convey the concept described in the ideal note, and is not commonly used in Spanish medical lexicon. This attribute represents the physician‟s skill to make a synthesis of the patient‟s status based on which a plan of care can be developed. “Interpretación” (interpretation) is the word that would reflect the concept in Argentina. The concept in English, generally speaking, is similar to “interpreting laboratory test results” where the information of laboratory tests helps doctors make decisions. Experts discussed this point and decided

that “interpretado” should be used in the consolidated version instead of “sintetizado” (synthesized).

The committee reviewed every description of the ideal notes and rearranged words and sentences as needed. Experts accepted the back translations of the title, even when they did not fully match the original version. The title of the consolidated version was the same as the one in the beta version: “Instrumento de Evaluación de la Calidad de la Documentación Médica”. There was not significant discussion about the Likert scale either. The final product was the consolidated version, created according to the suggestions submitted by the committee. Experts and back translators received the consolidated version by e-mail. Back translators agreed with the beta version. See Appendix 6: Evaluation of Backward Translations and Creation of the Consolidated Version Report.

Review Process

During the consolidation process, experts did seven reviews. They compared the beta version with the back translated versions, both back translated versions, the original version and the consolidated version, the original version and the back translated versions, and the consolidated version with the beta version. A matrix of agreement among versions was built. See Appendix 9: Comparison of the Original English Version with the Backward Translation, Appendix 10: Comparison of the Original English Version with the Beta and Consolidated Versions, and Appendix 11: Matrix of Agreement Among the Different Versions of the PDQI-9.)

1.5. Step 5: Pre-test

In this step, participants evaluated the clinical notes and suggested some changes to the format. The instrument was reduced to a one-page format and an appendix with additional information about the attributes was introduced. The average time for the entire evaluation of the clinical notes was 17.27 (SD 9.53, min 5, max 46) minutes. The average time for the evaluation of admission clinical notes was 13.78 (SD 5.25, min 5, max 25) minutes. The average time to evaluate progress notes was 15.07 (SD 7.42, min 8, max 34) minutes. The time for the evaluation of discharge summaries was in average 22.6 (SD 12.18, min 5, max 46) minutes.

Measurement of Readability of Ideal Notes Descriptions

Descriptions contained 284 syllables, 131 words, 10 phrases. The average syllables/words were 2.17, the average words/phrase was 13.10, and the Flesch-Szigriszt Index was 58.67. The word correlation was 12.18 and the Fernandez Huerta Index was 63.40. The PMOSE/IKIRSCH document readability result was 5 (combined list structure 2 points; fewer than 15 labels, 1 point; fewer than 75 items, 1 point; the reader received additional information about attributes, 1 point).

2. Phase 2: Validation of the Spanish Version of the PDQI-9 Score, Reliability and Validity Evaluation

The objective was to evaluate whether the translated and adapted instrument was valid and reliable. The following diagram (Figure 9: Phase two: Validation in the cross-

cultural adaptation process (adapted from (33))) depicts phase two workflow and the paragraphs below describe the process.

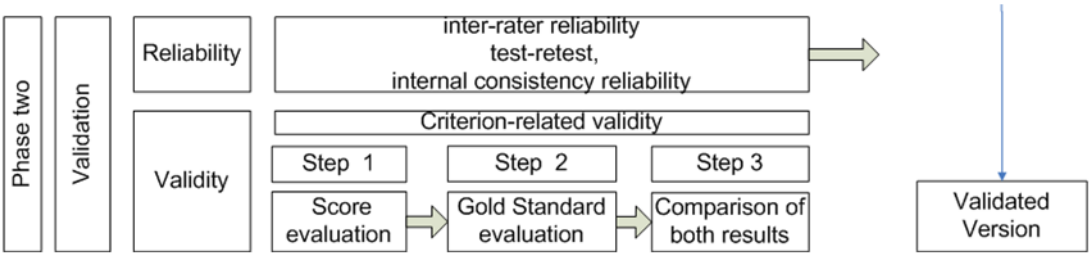


Figure 9: Phase two: Validation in the cross-cultural adaptation process (adapted from (33))

The implementation of phase two lasted from June to August of 2013. Seventy-four physicians were invited to participate (See Figure 10: Flowchart of study participant recruitment): sixty-one would use the adapted instrument (raters) and thirteen would not use the instrument (GS). Forty (63%) raters accepted to participate in the study. The main reason for not participating was lack of time to perform the evaluation. Twenty-five (43%) evaluated 30 clinical notes, and 10 (25%) re-evaluated 3 clinical notes. Twelve (92%) GS candidates accepted to participate in the study. Eight (62%) evaluated 30 clinical notes, and 3 (31%) re-evaluated the clinical notes designed for the retest.

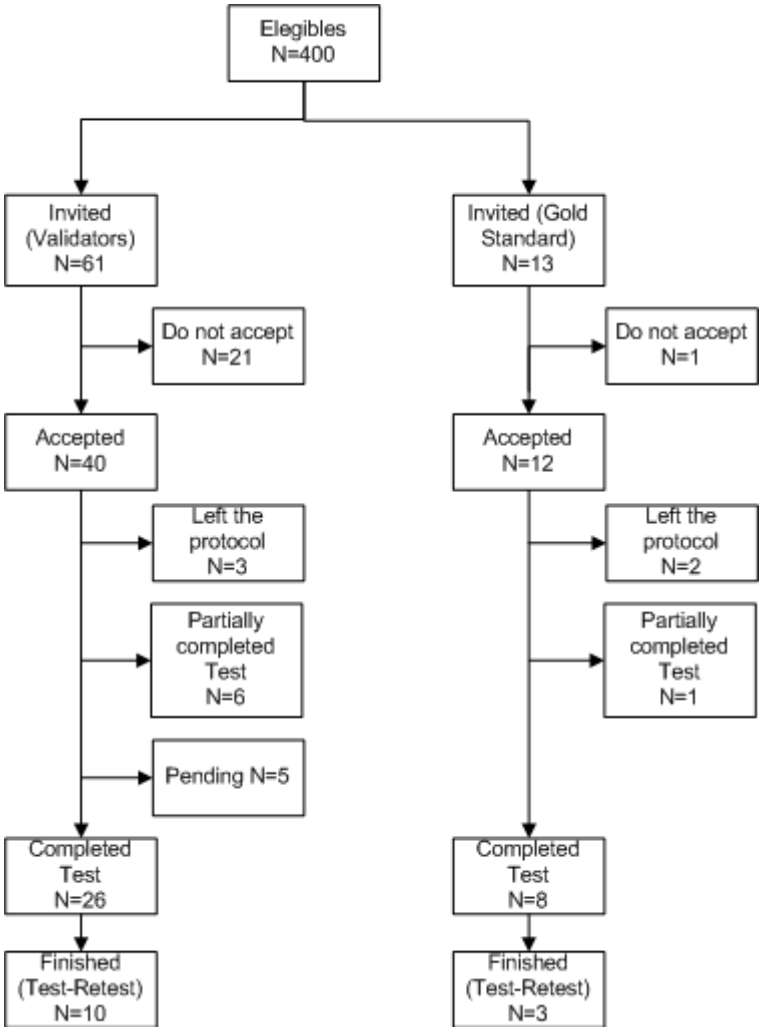


Figure 10: Flowchart of study participant recruitment

Participants included reviewed at least one clinical note. Forty physicians participated in the study, their mean age was 34 (SD 6.37). Twenty-one of them were female (52.5 %), the mean time from MD degree conferred was nine (SD 7.16) years. Seventeen (43%) physicians were specialists (or residents) in Internal Medicine and twenty-three (57%) had a second specialization in Emergentology (13%) , Infectology (10%), Critical Care (8 %), Endocrinology (5%), Gastroenterology (5%), Allergy (3%), Palliative Care (3%), Clinical

Pharmacology (3%), Rheumatology (3%), Pulmonology (3%), Hepathology (3%), and Hematology (3%).

Among **raters**, the mean age was 32 (SD 6.24). Nineteen of them were female (61%), and the mean time of MD degree conferred was 7.45 (SD 7.21) years. Fourteen (47%) were specialists in Internal Medicine and 17 (55%) had a second specialization in Emergentology (10%), Infectology (10%), Endocrinology (6%), Gastroenterology (6%), Allergy (3%), Palliative Care (3%), Clinical Pharmacology (3%), Rheumatology (3%), Pulmonology (3%), Hepathology (3%), and Hematology (3%). In the **GS** group, the average age was 39 (SD 3.54). Two (22%) were female, and the mean time since MD degree was awarded was 14 (SD 3.51) years. Three (33%) were specialists in Internal Medicine and six (66%) had a second specialization in Critical Care (33%), Emergentology (22%), and Infectology (11%). Table 4 summarizes the socio-demographic characteristics of participants.

Table 4: Socio-demographic characteristics of participants in the PDQI-9 Spanish version validation process

Socio-demographic characteristics (n 40)	
	%
Gender	
Female	47.50
Age (yr)	
25–29	25.00
30–39	55.00
> 40	20
Role	
Resident	35.00
Fellow	12.50
Staff	52.50
Specialty	
Subspecialties of Internal Medicine	57.50
n= number of participants yr= years	

The average time for the evaluation of the clinical notes was 10.28 (SD 6.81, min 1, max 70) minutes. The average time for evaluating admission clinical notes was 10.40 (SD 6.43, min 1, max 40) minutes. The average time for evaluating progress notes was 9.66 (SD 6.20, min 1, max 55) minutes. The average time for evaluating discharge summaries was 10.79 (SD 7.71, min 1, max 70) minutes. Some evaluations were performed in two different days, starting on day one and finishing the next day. For some of them the difference between day 1 and day 2 was negative. Because of this, they were eliminated from the analysis.

2.1. Reliability Evaluation

2.1.1. Evaluation of Internal Consistency (Cronbach's Alpha)

The overall Cronbach's Alpha coefficient was 0.92 for 898 evaluations. For admission notes (307 evaluations) it was 0.944, for progress notes (300 evaluations) it was 0.9. Finally, for discharge summaries (291 evaluations) the Cronbach‟s Alpha coefficient was 0.91.

2.1.2. Intra-rater Reliability (Test-retest) Evaluation of Intra-observer Variability (Reliability, Repeatability, Consistency)

The test-retest was performed in 30/898 (3.3%) clinical notes evaluated: 10 (33.3%) admission notes, 10 (33.3%) progress notes and 10 (33.3%) discharge summaries. The global mean score in the test was 34.3 (SD 5.6 CI 95 % 32.2 – 36.4) while in the retest was 32.9 (SD 5.6 CI 95% 30.8 – 34.95). The median of the differences between both scores was zero with a range from -5 to 11 with an interquartile range of 3. The mean score for admission notes was 34.7 (SD 5.3, CI 95% 30.9 – 38.5), the mean score in the retest was 34.8 (SD 5.3, CI 95% 31 – 38.6). The median of the differences between both scores was zero with a range from -5 to 3 with an interquartile range of 8. The mean score for progress notes was 31.6 (SD 5.5, CI 95% 27.6 – 35.6), the mean score in the retest was 30.3 (SD 4.8, CI 95% 26.9 – 33.7). The median of the differences between both scores was zero with a range from -2 to 9 with interquartile range of 5. The mean score for discharge summaries was 36.7 (SD 5.3, CI 95% 32.9 – 40.5), the mean score in the retest was 33.5 (SD 6.1, CI 95% 29.2 – 37.9). The median of the differences between both scores was zero with a

range from 0 to 11 with an interquartile range of 7. *The overall correlation coefficient was 0.79 (p <0.001).* The correlation coefficient for admission notes was 0.915 (p <0.001). The correlation coefficient for progress notes was 0.76 (p= 0.010). The correlation coefficient for discharge summaries was 0.72 (p= 0.018). The correlation coefficient for each dimension of the score is shown in the Table 5.

Table 5: Correlation coefficient of each score dimension

Clinical Notes (n)	**Correlation coefficient Test retest**	p value
Global (30)	0.79	<0.001
Admission notes (10)	0.91	<0.001
Progress notes (10)	0.76	0.010
Discharge summaries (10)	0.72	0.018
Attribute: Spanish (English)		
1. Actualizado (Up-to-date)	0.56	0.001
2. Preciso (Accurate)	0.47	0.009
3. Completo (Thorough)	0.65	<0.001
4. Útil (Useful)	0.61	<0.001
5. Organizado (Organized)	0.76	<0.001
6. Comprensible (Comprehensible)	0.5	0.004
7. Sucinto (Succinct)	0.75	<0.001
8. Interpretado (Synthesized)	0.68	<0.001
9. Internamente coherente (Internally Consistent)	0.56	0.001

Abbreviations: n, number of items

2.1.3. Inter-rater Reliability

Inter-rater reliability was measured using the Intra-class Correlation Coefficient (ICC) for consistency of average measures on the adapted PDQI-9 total scores for each of the 898 notes. In keeping with the original article, each note was considered as a fixed effect and each rater as a random effect, in a two-way mixed model (16). *The global ICC for 898 evaluations was 0.92 (95% CI 0.91 - 0.93, p<0,001).* For the admission notes (307) the coefficient was 0.94 (95% CI 0.93 -0.95, p<0,001), for progress notes (300) it was 0.90 (95% CI 0.88 - 0.92) and for discharge summaries (291) it was 0.91 (95% CI 0.89 –0.92).

2.2. Validity Evaluation

This step comprised the evaluation of the criterion validity of the adapted instrument.

2.2.1. Criterion Validity

The average and medians of the GS scores were estimated. The estimations for each note are shown the following figure (Figure 11: Average and medians of score evaluations by Gold Standards.). The figure shows two sources of variation: the experts and the documents themselves. There is a series of documents a through z, each with a score range represented with a box plot that illustrates (a) the score range for any single document –for most of them the range seems large– and (b) the score

range for the collection of documents. There are documents at the low end (for instance x and c) and documents at the high end (for instance ll and p) of the range.

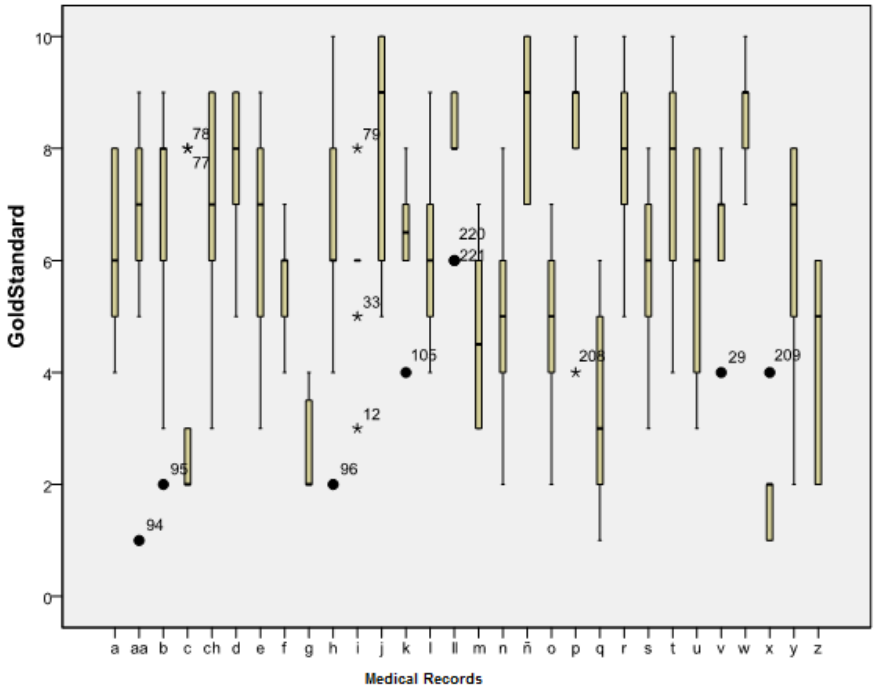


Figure 11: Average and medians of score evaluations by Gold Standards.

The records assessed were 898: 307 (34.2%) admission notes, 300 (33.4%) progress notes, and 291 (32.4%) discharge summaries. The correlation coefficient between the mean total score of raters and the average overall impression of the GS was 0.85 (p <0.001). The correlation coefficient between the mean total score of raters and the average overall impression by type of clinical note was, for admission notes, 0.97 (p <0.001), for progress notes, 0.84 (p 0.003), and for discharge summaries, 0.67 (p 0.035). See Appendix 12: Criterion Validity (Comparison with GSs)

Discussion

The aims of the study were to adapt to Spanish the PDQI-9 instrument and to validate the adapted version. The results of this study show that the adapted version of the PDQI-9 instrument is reliable and valid for it to be used in Spanish-speaking countries, for Internal Medicine physicians in institutions like Hospital Italiano de Buenos Aires. The adaptation process produced an adapted score, and three internal medicine physicians evaluated in the pre-test evaluation a reduced amount of clinical notes. At the end of the process, there were no changes in the score; the only changes were in the format of the form. In the validation process raters and GS Internal Medicine physicians performed evaluations on clinical notes. The instrument was found to present excellent internal consistency. Inter-rater reliability was very good and intra-rater reliability was good. Finally, there was a correlation between the instrument and the designed GS. The following sections contain a description of the psychometric measurements of the adapted instrument, a comparison between the latter and the original instrument, and a discussion about limitations, validity in mixed research and future implications.

Psychometric Measurements

The description and interpretation of the reliability and validity measurements of the adapted instrument, and a comparison with the original instrument are the focus of the sections below.

Reliability

The study shows that the adapted instrument has excellent overall internal consistency. The items in the adapted PDQI-9 instrument maintain an excellent degree of correlation in general and in particular for every type of clinical note, especially for admission notes and discharge summaries (> 0.90), while the degree of correlation was good for progress notes (0.90). In the original version, it was good for admission notes, and excellent for progress notes and discharge summaries. In both cases, it was excellent for discharge summaries. These results suggest that the score has a valid construct for the communication dimension. The alpha value changes depending on the population where the score is applied. That is the reason for evaluating internal consistency in the targeted population. The PDQI-9 instrument has good internal consistency, and so does its Spanish version, with an adequate value of Cronbach's Alpha coefficient. This means that items correlate well in the original population as well as in the targeted population. This test is valid only for one-dimensional domains, in this case, "communication". Knowing the internal consistency of an instrument is the first step in the process of validation against a GS and for some constructs it is the only kind of validation in the absence of a true GS (45).

As for inter-rater reliability, the Intraclass Correlation Coefficient (ICC) of the adapted version was higher than the ICC of the original version. The ICC of the Spanish version was very good (> 0.90) (46): there was almost perfect agreement (47). The ICC was good (46) for the original version, with almost perfect agreement (47). The intra-rater reliability (test-retest) measured for the Spanish version was good (> 0.80) (46), with substantial agreement (47). Inter-rater reliability refers to the agreement between two (or more) different raters when they evaluate the same measure in the same clinical note.

Variability among raters, in the instrument or the measuring process itself when it is performed at different times (re-test) can alter the correlation among measurements. Differences among raters, rather than in the measurement method, may explain the high degree of agreement observed (46). Therefore, in this case, the use of this instrument yields high results regardless of the raters in both populations. In the case of the adapted version, the higher ICC may be the result of the adaptation process, which led to a more stable and comprehensible instrument. Examples of the process are the cases of the attributes “up-to-date” and “useful”. In the case of the attribute “up-to-date”, the original version of the description of the ideal note contains the following text: "The note contains the most recent test results and recommendations." The final version added information, with the description of the ideal note reading as follows: "El registro contiene los resultados de los estudios, ***el estado clínico actual del paciente*** y recomendaciones más recientes." (“The note contains the most recent test results, recommendations, ***and the patient's current clinical status and latest recommendations.***") For the attribute “useful”, the description of the ideal note in the Spanish version was expanded too: "El registro es extremadamente relevante y útil, proporcionado información y/o análisis valiosos ***para el cuidado del paciente***." ("The note is extremely relevant, providing valuable information and/or analysis ***for patient care***.")

Criterion Validity

There was significant correlation between the average scores from the adapted instrument and the average GS overall impression. In terms of criterion-related validity, results are similar, although in the original study, the coefficient was measured between the General Impression and the PDQI-9 Total Score by comparing the evaluations for every note and, in this research, by comparing the average of evaluations. In the original study

correlation was from 0.678 to 0.838 (p= 0.000) while in the present study it was 0.85 (p <0.001).

Limitations

This study showed that the adapted instrument is reliable and valid; however, some limitations should be taken into consideration. Some of them are related to the original research (the use of printed copies of the clinical notes and the physicians‟ perspective at the time of evaluating), while others are inherent to the selection of the participants and the clinical notes, the type of study, the administration of the score, and the way of grading clinical notes. The paragraphs below discuss such aspects.

First, the study used pdf printed copies of clinical notes. The simulated conditions may affect the evaluation of the quality of clinical notes and, hence, the evaluation results. The original research used hard copies of the clinical notes. In this study, it would have been necessary to print out about a thousand hard copies and, therefore, the logistics of de-identification of the clinical notes and paper storage would have rendered the project impractical. This simulated condition, however, allowed evaluation portability. By having pdf copies available, physicians were able to access clinical notes when they needed to perform the evaluations, working at their own pace, even from outside the hospital, without being constrained by connectivity issues or possible downtimes of the system during the evaluation. This represented an advantage. All they needed was a computer (39), and many physicians would be able to access the information at the same time (48). However, making information so available posed some challenges too. The simulated environment could affect the manner in which physicians performed the evaluations. Supplementary research aimed to

assess performance with the real environment and a usual user interface is necessary to learn more about the possible impact on the quality of evaluation.

Second, this original research based its conclusions on the physicians‟ perspective of the quality of clinical notes, without taking into account other clinicians or patients‟ viewpoints. Internal Medicine physicians‟ perspective on the quality of clinical notes may be different from that of other specialties, and the original research pointed out this limitation. The standpoint of other physicians, clinicians or patients is important and can complete all the aspects of the quality of clinical notes. However, focusing on one group of individuals may help understand better what the characteristics of the quality of clinical notes are. These aspects were addressed by Hanson et al. (27) when they evaluated the quality of clinical notes.

Third, the PDQI-9 instrument sets the attributes considered important for proper communication among physicians when they evaluate clinical notes. While it enables this important purpose, there are other aspects such as billing, educational, research, or legal reasons that are not evaluated. Focusing on communication, however, makes it easier to evaluate clinical notes with a clear purpose.

Besides the above mentioned limitations, there are some other constraints that are intrinsic to this study. *Fourth*, this study used non-probability sampling (convenience and snowball sampling), which may introduce sampling bias. However, a random sampling would turn the project unfeasible (49). In an attempt to reduce bias, the master‟s candidate avoided selecting any participant by herself. To select experts at least four Internal Medicine physicians were consulted besides the master‟s candidate (one member of the thesis committee of the master‟s candidate, one chief of residence in Internal Medicine and two

key representative physicians of the Internal Medicine Department). Physicians who helped in the selection did not participate in any phase of the study.

Fifth, only one physician selected the clinical notes for this study. The selection was based on his quality criteria. These criteria may not match the criteria of the rest of physicians who acted as GS in the evaluations. In addition, in the selection of the medical records, the patient‟s length of stay at the hospital was not limited and, hence, some medical records had many pages to evaluate (one hundred or more), turning the review process cumbersome and difficult. On the other hand, limiting by the patient length of stay would not reflect the population of clinical notes under study. Additional research may consider two or more physicians in the selection process of clinical notes and judge the convenience of using medical records of patients with shorter hospitalizations.

Other limitations are specific of this type or research. *Sixth*, in cross-cultural studies, in the adaptation phase, a qualitative analysis may be insufficient for assuring content validity. Some authors suggest conducting quantitative analysis (50), in addition to the qualitative analysis, as a means to enhance content validity. Conversely, the quantitative analysis is an instance of the agreement, like a snapshot in the process. The qualitative analysis procedure per se allows exploring the complexity of representing the concept in the targeted culture. If experts “share common elements in the understanding of the same culture” that is enough for the content to be valid (51).

Seventh, it is recommended that professional translators, professional back translators, and a bilingual/multilingual expert committee contribute in this type of research (51). Unfortunately, this kind of requirements may limit the process and render a project impracticable. On the other hand, it is suggested that “qualified” translators/back translators

(17) participate in the process instead of "professional" translators. They are "qualified" because they "culturally" represent the target population. The translators/back translators in this study knew the domain and the representation of the concept applied in the local culture. That knowledge made them "qualified" translators.

Eighth, subjective phenomena such as the evaluation of quality can be considered "soft" measures (52). In the validation of tests (including the validation phase of the cross-cultural studies), there are concerns about the "true gold standards" or even reference standards (52). In this study, Internal Medicine physicians considered experts acted as GS because there is not a "true gold standard" for clinical notes. Physicians followed their own criteria for evaluating clinical quality notes. This "imperfect reference standard" (53) had high variability at the moment of evaluation. Further research can improve the reliability of the "imperfect reference standard", for example, a) by eliminating the most controversial evaluations through statistical procedures, or b) by homogenizing the concepts shared among the GS, training them, and performing a second quality evaluation.

In addition to the limitations already mentioned, there are some others related to score administration. The limitations as to how the form was administered were similar to those of the self-administered instruments of measurement. In self-administered instruments, the understanding of the instrument affects results (54). *Ninth*, because there might be less interaction between the researcher and evaluators (39), the evaluation may be incomplete or inaccurate (54-55) due to distractions when doing the evaluation (56) or the absence of the researcher to explain the doubts when the evaluator performed the assessments (54-56). In this study, e-mail, phone contacts, and face-to-face communication, attempted to overcome this limitation on interactions between raters/researcher and between GS/researcher. Other

than this disadvantage, less interaction between the researcher and the evaluators reduces the possibility of researcher bias (54).

Tenth, another score administration-related drawback is the lack of direct control of physicians‟ actions (54-57). In this study, physicians were instructed to conduct the evaluations individually. However, there was no control over whether physicians consulted or discussed with other physicians about the quality of specific clinical notes. The honesty and integrity of physicians were taken for granted. Reinforcing this assumption, and in an attempt to reduce the possibility of cross-information physicians were randomly assigned to different groups.

Eleventh, following with the drawbacks of self-administered instruments, they may have a low response rate (55). Raters‟ characteristics or motivations may differ, thus producing a non-response bias (39, 54). In this study, thirty percent of raters/GS did the assessment within the originally scheduled time. Many physicians reported lack of time to do the evaluations. To increase the rate of response, the physicians were given the possibility of turning over the forms as they completed them, even if they had not done all thirty evaluations and the length of the evaluation was extended. This contributed to increasing interaction with participants and creating a feeling of relief in some of them, as they had a sense of task accomplishment. To increase the likelihood of completing the sample size a second group of physicians were invited to join in. In addition, individuals received 3 reminders by e-mail as suggested in the literature (58). It is worth mentioning that, as "questionnaires are simply not suited for some people"(57), the evaluation of clinical notes using a score is a job that not all physicians are prepared to do. Some of them mentioned that the task of quality evaluation is more appropriate for medical auditors.

Physicians who did not do the evaluations in the first weeks tended to do it with the hard copies in the following weeks. Most physicians asked for a hard copy of forms with the score. Even using a Word document to fill out the form, not having a web-based instrument added advantages and disadvantages. The disadvantage of not using a web- based instrument was that data had to be entered in a database. The transcribing process was time consuming and not free of errors (55). Some physicians reported missing hard copies. They received a new printed set of forms. *Twelfth*, the last score administration-related limitation, information being gathered from a self-administered instrument responds to predetermined items (48, 57).For that reason, to enhance the flexibility of the instrument, a space for comments was added (57). The content analysis of these comments is still pending.

Other limitations may be related to the way in which clinical notes are graded. In the education field, the bias in grading can be direct or indirect, conscious or not, and positive or negative (59). There is evidence that some factors may influence educators in their grading and assessment practices. These factors are aspects of the evaluator, the student, the environment, the characteristics of the course, the type of activity, the type and quality of the paper being evaluated, and the order of the evaluation (60-63).

Thirteenth, in this thesis, aspects of the evaluator (raters/GS), the environment, and the type of activity, the type and quality of the clinical note under evaluation, and the order of the evaluation are important aspects to consider. Among the aspects that are connected to the evaluator, the specialty/subspecialty, daily workload (full or part-time), her/his tendency to be indulgent or not, and if he/she is resident or attending may influence grading. Repetitive or monotonous activities may affect as well the way in which an evaluator grades a paper. Sometimes, the quality of written papers has an effect on rating; the evaluator may

overrate or underrate documents depending on the order of the evaluation. The amount of patients (clinical notes and medical records) to evaluate and hospital size may impact on grading. Double de-identification (patient and physician who created the clinical note) contributed to preventing physicians from being influenced at the time of grading the clinical notes.

Finally, these results cannot be generalized. Three factors can affect external validity: the people who participated in the study, the setting of the study, and the time when the study was conducted (64). Physicians may have a different workload or workflow, as compared to those of other physicians in other institutions. In addition to that, the characteristics of HIBA as a private hospital, or the time when this study was done, when new medical residents were beginning their residency, may have influenced the results of this research. Therefore, it is recommended that an additional cross-cultural adaptation and validation of the PDQI-9 instrument in other Spanish speaking countries healthcare institutions be conducted. Despite the limitations, this study presented a reliable and valid nine-attribute instrument for the assessment of quality of clinical notes. This instrument can be used as a basis for additional research.

Validity in Mixed Research

Cross-cultural validation takes into account a similar phenomenon in different cultures for the purpose of obtaining an equivalent instrument. It considers the cultural context, idioms, and the different perspectives on the observable fact so that the adapted instrument fits in the targeted population (17, 33). This multi-step research has a mixed research approach. Mixed method research "combines elements of qualitative and quantitative research approaches for the broad purposes of breadth and depth of

understanding and corroboration" (65). In the "qualitative-quantitative continuum" cross-cultural validation would correspond to the "equal status" (QUAL-QUANT) where the qualitative-quantitative approach considers the research questions (65). Some studies have proposed the use of the term "legitimation" instead of "validity" in mixed research. The concept of "validity" applies to quantitative research and "transferability" to qualitative research. "Legitimation" would involve both methods (66). Among the types of legitimations proposed, the "multiple validities legitimation" would apply in this case. Cross-cultural validations address and accomplish the integration of quantitative and qualitative "validities" as a whole (66), because both parts are important in the process. In keeping with this reasoning, it is important to expand some concepts on "equivalence" and the "validity of the instrument" to see the degree of legitimation obtained.

Equivalence in the Adaptation Process

The use of a score in a different setting, country, language or culture determines the necessity of conducting a cross-cultural adaptation beyond a simple translation, because the objective of the cross-cultural adaptation is to ensure equivalence with the original instrument (17). In the following section, the discussion will focus on the achievement of equivalence in the process.

It has been mentioned that the objective of cross-cultural adaptation is to have equivalence, specially semantic, idiomatic, experiential and conceptual equivalence (17). There are many types of equivalence, such as functional, technical, scalar/metric, scale, operational, item, cultural, language, construct, measurement equivalence, in addition to the above-mentioned type of equivalence. Sometimes there is no consensus about the definition of the different types of equivalence; or their descriptions overlap (67). The meaning of

words, or **semantic equivalence,** may depend on the context. In order to achieve semantic equivalence in this research, the vocabulary and grammar of the translated score were modified in the process of expert review. **Idiomatic equivalence** is the equivalence of idioms or colloquialisms between the original version and the target culture. The expert panel checked this and there were idioms neither in the original version nor in the Spanish version.

In **experiential equivalence**, the circumstances or conditions of the target cultural context are similar to those of the original cultural context (67). The condition of evaluating inpatient clinical notes at the Internal Medicine Department in an academic hospital fits in the target population. The physicians at the pre-test evaluation completed the evaluation; they understood the concepts and suggested changes in the format, not in the content, of the instrument. In addition, readability tests showed that the ideal description of the adapted instrument was “normal”, that means the text to be accessible to ordinary people (>= 60). In terms of complexity evaluation, the document has a "very low" complexity level (for grade 4 or less than 8 years of schooling).

The results obtained in this study are consistent with the findings of the original research. In the original study, participants spent 10 minutes or less to review the clinical note and score it, spending up to 90 minutes to review all charts and score all nine clinical notes (16). In this project, the average time to complete the evaluation in the pretest was 17 minutes, while the validation process took 10 minutes. Five of the nine attributes of the PDQI-9 score (organized, comprehensible, succinct, synthesized, and internally consistent) represent attributes of the evaluation of the text that do not depend on the contextual information of the medical record. However, attributes like “up-to-date”, “accurate”, or

"thorough" rely on supporting information that must necessarily be read to understand the patient‟s status; this might explain why evaluating clinical notes may take time.

Conceptual equivalence is accomplished when the concept exists and is valid in the targeted culture (17). Semantic and concept equivalence are related. The concept and the expression may exist or not in the original and the target culture, but at least one of them, concept or expression, are required to recreate the equivalence in the target language (68). All of the attributes and concepts in the PDQI-9 score in the original language exist in Spanish, as do the words and concepts in the description of the ideal note, the instructions and the title. Other important types of equivalence are **content equivalence** and **criterion equivalence** (67). In **content equivalence,** the content of the concept is important/relevant to the targeted culture (67). To ensure that the consolidated version maintained content validity, the aim was to express a concept in both languages and assess its relevance in the medical culture (17). Through a qualitative analysis, the panel of experts made adjustments to the wording of the items, the description of ideal notes and the title to ensure that the Spanish version of the PDQI-9 instrument would be applicable within a Spanish-speaking context. In **criterion equivalence** there is a relationship between the adapted instrument and independent criteria of the targeted culture. This type of equivalence can be achieved via psychometrics measurements. Having the instrument adapted, the next objective is to evaluate whether the instrument is reliable and valid.

Instrument Validity

According to classic theories, there would be different types of validity: "face validity", "content validity", "criterion validity", and "construct validity". Trochim (64) grouped face and content validity as "translation validity" procedures, regardless of whether

the instrument is going to be used in another language or not. In assessing **face validity**, experts evaluate whether the attributes in the instrument look like or have the characteristics of interest (69). **Content validity** is assessed by qualitative methods when experts evaluate whether the instrument represents the phenomena of interest (70-71). **Criterion validity** measures the agreement between the instrument and other criterion valid in the target population (68). **Construct validity** assesses the performance of the instrument in practice (70). **Convergent or divergent validities** are part of the evaluation of construct validity. There is convergent validity when measures of expected related constructs are related when they are observed (64).

Other researchers conceptualize these types of validity under a more global concept (64, 72). "(...) Construct validity is related to generalizing (...) involves generalizing from (one) program or measures to the concept of (other) program or measures..." (64) There is "construct validity" when the interpretations (inferences) done with the instrument represent the construct and there is enough evidence that supports the relationship between the inferences and the construct. The basis of evidence of "construct validity" is given by five possible sources: the "content", the "response process", the "relation to other variables", the "internal structure" and the "consequences" (73). According to Cronbach, the researcher should pursue construct validity through a "construct validation procedure" (71). In fact, "construct validity" is a matter of getting enough proof by performing translation validity procedures, pretest evaluations, criterion validity measurements, reliability measurements, convergent/divergent validity measurements and evaluating the possible consequences of using the instrument (72). This concept applies for selecting, adapting or developing an instrument and is depicted in the following figure (Figure 12: The construct validity

framework (69, 72-73)) showing the relations among source of evidence, measurements and types of equivalence.

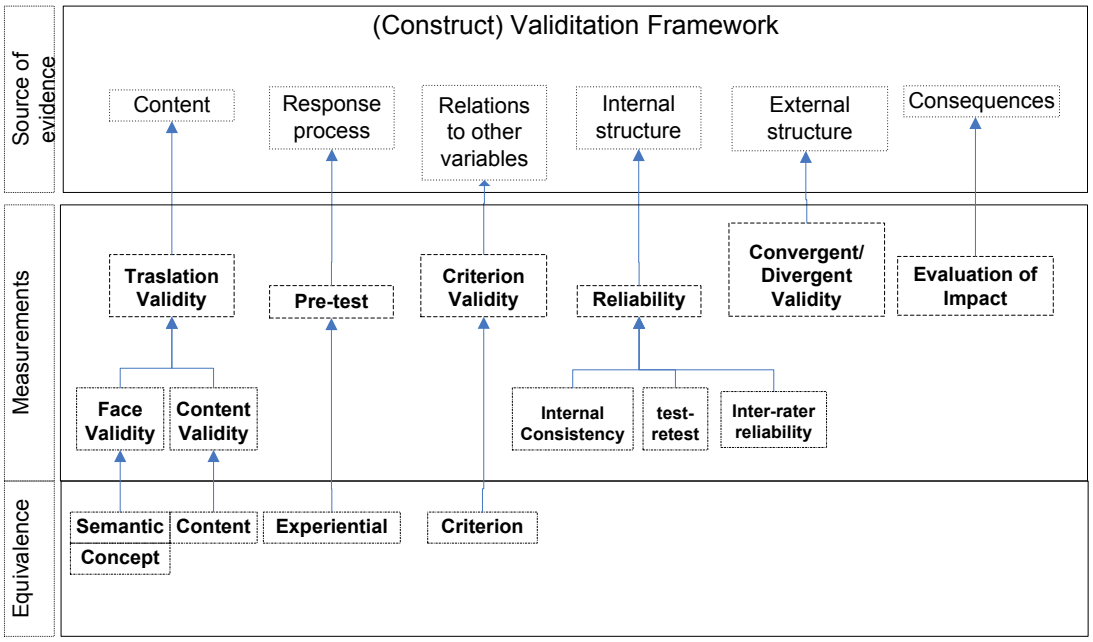


Figure 12: The construct validity framework (69, 72-73)

In summary, for the adapted version, "legitimation" or "validity" was obtained by combining different sources of evidence that support the construct validity. They are "content" and "response process" for phase one, and "relations to other variables" and "internal structure" for phase two. Time and enough amounts of clinical notes evaluated are required for supporting (or not) "external structure" and "consequences".

Implications/Recommendations for Practice and Future Research

This section provides the implications of the findings accomplished as well as recommendations for further research.

The implications of this study are grouped into academic and practical implications. From an academic point of view, this study: 1) helps to fill the knowledge gap in the evaluation of the quality of clinical notes in Spanish speaking countries, expanding the understanding obtained from the original project, 2) performs an integration of the literature reviewed and presents an adapted framework of the construct validity that combines equivalence, source of evidence, and measurements. This framework can be used for any instrument other than the cross-cultural validation study.

The recommendation for researchers from another language, who are interested in using an instrument like this, is to follow the two-step methodology to get a valid and reliable instrument. The key point is to have an equivalent instrument and have the essential construct structured in the same way across different cultural groups. Diverse forms of bias may influence the equivalence, especially the test bias. The test bias groups the construct, method and the item bias (74). The two-step framework standardizes the strategies suggested to reduce this bias (74). The equivalence is achieved mostly in the translation-adaptation process, which is the first step, but it is incomplete if is not compared with an external criteria, that is accomplished in the second step.

From a practical standpoint, the findings of this research 1) will allow the evaluation of the quality of clinical notes at HIBA, and other similar organizations, and 2) would help to improve the continuity of care. The standardization process of the quality evaluation will

permit interventions to improve the quality and indirectly would influence on the quality of care while helping in its continuity. There is, therefore, a definitive necessity for adapting and validating the instrument in other specialties or areas like surgery, pediatrics, orthopedics, nursing, kinesiology, or other domains such as imaging. In turn, this research may help healthcare organizations such as those similar to HIBA, as well as other healthcare organizations.

The research in this area is still in infancy. Researchers interested in the quality of clinical notes should be aware that clinical notes have "expressivity", which originates in a combination of narrative and structured text (75) in their creation. This circumstance imposes complexity in the evaluation of the quality of clinical notes. In addition, while increasing the challenge further, there is an absence of a true Gold Standard. On the other hand, the perspective of quality evaluated with this score is "communication" for continuity of care. Further research is necessary to answer if the attributes in the PDQI9 score are sufficient in order to evaluate the quality of clinical notes in Internal Medicine or other specialties. That is, more evaluations will answer if the construct of the score is valid. Beyond this perspective, it would be advisable to evaluate if the condition of good quality communication will help to teach and to research, in a legal or financial process. Furthermore, if a clinical note "communicates" with quality, how that communication improves patient care.

Conclusion

This section provides an overview of the research and the conclusions.

The narrative components of medical records, especially admission notes, progress notes and discharge summaries, become important for communication between health care providers in the process of the continuity of patient care. The increasing implementation of EHRs and the use of functionalities have shown unintended consequences: poor quality of clinical notes characterized by "unreadable" notes with redundant information. This instance of miscommunication clearly can affect the continuity of patient care. In this scenario, it is critical to evaluate the quality of clinical notes. The focus on the evaluation of the quality of medical records has been on attributes like completeness and accuracy. For clinical notes, however, the expansion to other domains of evaluation is indispensable. PDQI9 is one of the instruments recently developed that evaluates the attributes related to the quality of clinical notes. The instrument focus, unlike others, is on the communication between Internal Medicine physicians.

Knowing the importance of the topic and the advantages of the PDQI9 score, this Master project proposed to make a Spanish version of the PDQI9 through a cross-cultural adaptation and validation process. Following the two-step process ensures getting an equivalent instrument in the target population and reduces the test bias. The process of the cross-cultural study was carried out at HIBA, in Buenos Aires Argentina. The first phase was the translation and cross-cultural adaptation of the instrument. The achievement of the content validity (among others) was performed through a review of a panel of experts. The second phase evaluated the reliability and validity of the Spanish version. The

accomplishment of the criterion validity was performed through the comparison of the results from the adapted instrument with the results obtained from the experts‟ evaluation (Gold Standards for this study).

After this two-step process, the instrument was adapted, reliable, and valid. However, because of the cultural differences with other Spanish speaking countries, the recommendation is to perform an extra cultural adaptation and validation of the new instrument in those countries. Among the limitations described in this research, one of the most important is that there is not a true Gold Standard. There was a high variability in the performance of the Internal Medicine expert physicians when they rated the clinical notes. Standardization in the way physicians create and evaluate the quality of clinical notes can reduce this variability.

The academic implication of this study is the evaluation of the construct validity across time, not only using a statistical test, but also as a whole concept where there is a combination and construction of the validity (or legitimation) with the different types of equivalence, source of evidence, and measurements. From a practical point of view, the findings of this research will help in the process of evaluation of the quality of clinical notes and indirectly, influence in the continuity of care.

In conclusion, from this research, the cross-culturally adapted Spanish version of the PDQI9 instrument is reliable and valid for Internal Medicine physicians, for its use in Spanish-speaking countries, especially in Argentina at the “Hospital Italiano de Buenos Aires”. Further research is needed to answer if the attributes included in the score are sufficient to determine the quality of clinical notes and if a clinical note with “good quality”

in communication has the same weight for teaching and researching, in a legal or financial process, in the same or other specialties or healthcare professions.

References

1. Embi PJ, Weir C, Efthimiadis EN, Thielke SM, Hedeen AN, Hammond KW. Computerized provider documentation: findings and implications of a multisite study of clinicians and administrators. J Am Med Inform Assoc. 2013 Jan 25.

2. Weir CR, Hammond KW, Embi PJ, Efthimiadis EN, Thielke SM, Hedeen AN. An exploration of the impact of computerized patient documentation on clinical collaboration. Int J Med Inform. 2011 Aug;80(8):e62-71.

3. Sutcliffe KM, Lewton E, Rosenthal MM. Communication failures: an insidious contributor to medical mishaps. Acad Med. 2004 Feb;79(2):186-94.

4. Agency for Healthcare Research and Quality PSNet. Patient Safety Primers: What are Patient Safety Primers? Handoffs and Signouts. 2012 [cited 2013 02/15]; Available from: http://www.psnet.ahrq.gov/primer.aspx?primerID=9.

5. JCAHO. The Sentinel Event Data Root Causes by Event Type, 2004- 2012. 2013; Available from: http://www.jointcommission.org/assets/1/18/Root_Causes_Event_Type_04_4Q2012.pdf.

6. Amatayakul M. Definition of the Health Record for Legal Purposes (AHIMA Practice Brief). Journal of AHIMA. 2001;72(9):88A-H.

7. HIMSS. EHR: Electronic Health Record. [cited 2012 September, 5]; Available from: http://www.himss.org/ASP/topics_ehr.asp.

8. Quirós F, Luna, D. Electronic health records. In: SEIS-ECLAC, editor. eHealth handbook for managers of healthcare services and systems:. Madrid: United Nations Publication; 2012. p. 57-72.

9. Warner D, Wiedemann LA. Managing copy functionality and information integrity in the EHR. J AHIMA. 2012 Mar;83(3):47-9.

10. Embi PJ, Yackel TR, Logan JR, Bowen JL, Cooney TG, Gorman PN. Impacts of computerized physician documentation in a teaching hospital: perceptions of faculty and resident physicians. J Am Med Inform Assoc. 2004 Jul-Aug;11(4):300-9.

11. Weir CR, Nebeker JR. Critical issues in an electronic documentation system. AMIA Annu Symp Proc. 2007:786-90.

12. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc. 2004 Mar-Apr;11(2):104-12.

13. Beasley JW, Wetterneck TB, Temte J, Lapin JA, Smith P, Rivera-Rodriguez AJ, et al. Information chaos in primary care: implications for physician performance and patient safety. J Am Board Fam Med. 2011 Nov-Dec;24(6):745-51.

14. Committee on Quality of Health Care in America IoM. To Err Is Human: Building a Safer Health System2001.

15. Stetson PD, Morrison FP, Bakken S, Johnson SB. Preliminary development of the physician documentation quality instrument. J Am Med Inform Assoc. 2008 Jul-Aug;15(4):534-41.

16. Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing Electronic Note Quality Using the Physician Documentation Quality Instrument (PDQI-9). Appl Clin Inform. 2012;3(2):164-74.

17. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol. 1993 Dec;46(12):1417-32.

18. Gjersing L, Caplehorn JR, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. BMC Med Res Methodol. 2010;10:13.

19. Reichenheim ME, Moraes CL. [Operationalizing the cross-cultural adaptation of epidemiological measurement instruments]. Rev Saude Publica. 2007 Aug;41(4):665-73.

20. Carvajal A, Centeno C, Watson R, Martinez M, Sanz Rubiales A. ¿Como validar un instrumento de medida de la salud? Anales del Sistema Sanitario de Navarra. 2011;34:63-72.

21. Arribas A. Adaptación transcultural de instrumentos, guía para el proceso de validación de instrumentos tipo encuestas. Revista Científica de la AMBB. 2006;16(3):74-8.

22. Juran J, Godfrey, A Juran's Quality Handbook . 5th edition. McGraw-Hill, editor1999.

23. English LP. Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits1999.

24. American Society of Quality - Quality Glossary - Q. [cited 2012 11/27]; Available from: http://asq.org/glossary/q.html.

25. Logan JR, Gorman PN, Middleton B. Measuring the quality of medical records: a method for comparing completeness and correctness of clinical encounter data. Proc AMIA Symp. 2001:408-12.

26. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997 Sep-Oct;4(5):342-55.

27. Hanson JL, Stephens MB, Pangaro LN, Gimbel RW. Quality of outpatient clinical notes: a stakeholder definition derived through qualitative research. BMC Health Serv Res. 2012 Nov 19;12(1):407.

28. Hammond KW, Efthimiadis EN, Weir CR, Embi PJ, Thielke SM, Laundry RM, et al. Initial Steps toward Validating and Measuring the Quality of Computerized Provider Documentation. AMIA Annu Symp Proc. 2010;2010:271-5.

29. Shen S, South BR, Butler J, Barrus R, Weir C. The relationship between structural characteristics of 2010 challenge documents and ratings of document quality. AMIA Annu Symp Proc. 2012;2012:848-55.

30. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975 Nov;12(3):189-98.

31. Lobo A, Ezquerra J, Gomez Burgada F, Sala JM, Seva Diaz A. [Cognocitive mini-test (a simple practical test to detect intellectual changes in medical patients)]. Actas Luso Esp Neurol Psiquiatr Cienc Afines. 1979 May-Jun;7(3):189-202.

32. Fernández A, Pérez E, Alderete A, Richaud M, Fernández Liporace M. ¿Construir o Adaptar Tests Psicológicos? Diferentes Respuestas a una Cuestión Controvertida. Evaluar. 2011;10 60 - 74.

33. Ramada-Rodilla JM, Serra-Pujadas C, Delclos-Clanchet GL. [Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations]. Salud Publica Mex. 2013 Feb;55(1):57-66.

34. Ley Nacional 25 326: Ley de Protección de los Datos Personales [Argentine National Law on Personal Data Protection]. Sancionada por el Senado y Cámara de Diputados de la Nación Argentina reunidos en Congreso; 2000 [updated 2008; cited 2013 03/22]; Available from: http://www1.hcdn.gov.ar/dependencias/dip/textos%20actualizados/25326.010408.pdf.

35. Barrio-Cantalejo IM, Simon-Lorda P, Melguizo M, Escalona I, Marijuan MI, Hernando P. [Validation of the INFLESZ scale to evaluate readability of texts aimed at the patient]. An Sist Sanit Navar. 2008 May-Aug;31(2):135-52.

36. Gremy F, Degoulet P. Assessment of health information technology: which questions for which systems? Proposal for a taxonomy. Med Inform (Lond). 1993 Jul-Sep;18(3):185-93.

37. Gröne Oana. Inventario de instrumentos para medir la legibilidad de un texto. Grupo de health literacy , Red Catalana de Hospitales Promotores de la Salud [serial on the Internet]. 2010: Available from: http://www.udg.edu/LinkClick.aspx?fileticket=7Uf8HjTW-IU%3D&tabid=15493&language=ca-ES.

38. Mosenthal PB, Kirsch IS. A new measure for assessing document complexity: The PMOSE/IKIRSCH document readability formula. Journal of Adolescent & Adult Literacy. 1998;41(8):638-57.

39. Mitchell ML, Jolley JM. Research Design Explained, 8th ed: Wadsworth; 2012.

40. The PMOSE/IKIRSCH Document Readability Formula: A tool for assessing documents. 2012 [cited 2013 08/13]; Available from: http://www.hsph.harvard.edu/healthliteracy/files/2012/09/pmose.pdf.

41. Acerca de las estructuras educativas y del sistema de calificación y promoción. Buenos Aires2011 [cited 2013 08/11]; Available from: http://portales.educacion.gov.ar/vnt/acerca-de-la-estructura-educativa/.

42. Antunes B, Daveson B, Ramsenthaler C, Benalia H, Ferreira PL. The Palliative care Outcome Scale (POS) Manual for cross-cultural adaptation and psychometric validation2012: Available from: http://pos-pal.org/doct/Manual_for_crosscultural_adaptation_and_psychometric_validation_of_the_POS.pdf.

43. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011 Jan;64(1):96-106.

44. Muñiz J, Elosua P, Hambleton RK. Directrices para la traduccion y adaptacion de los tests: segunda edicion. Psicothema. 2013;25(2):151-7.

45. Campo-Arias A, Oviedo HC. Revision/Review Propiedades Psicometricas de una Escala: la Consistencia Interna. Revista de Salud Publica. 2008;10:831-9.

46. Pita Fernandez S, Pertegas Diaz S. La fiabilidad de las mediciones clinicas: el analisis de concordancia para variables numericas. Cad Aten Primaria. 2003;10(4):290-96.

47. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-74.

48. Cargan L. Doing Social Research: Rowman & Littlefield Publishers; 2007.

49. Laerd d. Non-probability sampling. 2012; Available from: http://dissertation.laerd.com/non-probability-sampling.php#step2.

50. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. Western Journal of Nursing Research. 2003;25(5):508-18.

51. Chavez LM, Canino G. Toolkit on translating and adapting instruments. Human Services [serial on the Internet]. 2005: Available from: http://www.hsri.org/files/uploads/publications/PN54_Translating_and_Adapting.pdf.

52. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. BMJ. 2002 Feb 23;324(7335):477-80.

53. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol Assess. 2007 Dec;11(50):iii, ix-51.

54. Passmore C, Dobbie AE, Parchman M, Tysinger J. Guidelines for constructing a survey. Fam Med. 2002 Apr;34(4):281-6.

55. Jones S, Murphy F, Edwards M, James J. Doing things differently: advantages and disadvantages of Web questionnaires. Nurse researcher. 2007;15(4):15-26.

56. Lewis-Beck MS, Bryman AE, Liao TF. The SAGE Encyclopedia of Social Science Research Methods: SAGE Publications; 2004.

57. Walonick D. Designing Surveys and Questionnaires. 2013 [cited 2013 08/23]; Available from: http://www.statpac.com/surveys/disadvantages.htm.

58. Burns KE, Duffett M, Kho ME, Meade MO, Adhikari NK, Sinuff T, et al. A guide for the design and conduct of self-administered surveys of clinicians. CMAJ. 2008 Jul 29;179(3):245-52.

59. Malouff J. Bias in Grading. College Teaching. 2008 2013/08/27;56(3):191-2.

60. Klein J. The failure of a decision support system: inconsistency in test grading by teachers. Teaching and Teacher Education. 2002;18(8):1023-33.

61. Dingfelder S. What's in a grade? [cited 2013 08/27]; Available from: http://www.apa.org/gradpsych/2006/03/cover-grade.aspx.

62. Hu S. Correlates of College Grades. Beyond grade inflation: Grading problems in higher education: Jossey-Bass; 2005.

63. Duncan CR, Noonan B. Factors affecting teachers' grading and assessment practices. Alberta Journal of Educational Research. 2007;53(1).

64. Trochim WM. The Research Methods Knowledge Base. 2005 [cited 2013 08/18]; Available from: http://www.socialresearchmethods.net/kb/.

65. Johnson RB, Onwuegbuzie AJ, Turner LA. Toward a definition of mixed methods research. Journal of mixed methods research. 2007;1(2):112-33.

66. Onwuegbuzie AJ, Johnson RB. The validity issue in mixed research. 2006.

67. M. Herdman SF-RJBX. 'Equivalence' and the translation and adaptation of health-related quality of life questionnaires. Quality of Life Research: Springer Science & Business Media B.V.; 1997. p. 237-47.

68. Orley J, Kuyken W, Patrick DL, Wild DJ, Johnson ES, Wagner TH, et al. Cross-Cultural Validation of Quality of Life Measures. Quality of Life Assessment: International Perspectives: Springer Berlin Heidelberg; 1994. p. 19-32.

69. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. 4 - Planning the Measurements: Precision and Accuracy. Designing Clinical Research. 3rd ed: Lippincott Williams & Wilkins; 2007.

70. Glossary of statistical terms. The Institute for Statistics Education; 2004-2013 [cited 2013 08/19]; Available from: http://www.statistics.com/glossary/.

71. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychological bulletin. 1955;52(4):281.

72. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. The American journal of medicine. 2006;119(2):166. e7-. e16.

73. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American psychologist. 1995;50(9):741.

74. de Klerk G. Cross-cultural testing. M. Born, C.D. Foxcroft & R. Butter (Eds.), Online Readings in Testing and Assessment; 2008 [cited 2013 12/01]; Available from: http://www.intestcom.org/Publications/ORTA/cross+cultural+testing.php.

75. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc. 2011 Mar-Apr;18(2):181-6.

76. Stetson PD, Bakken S, Wrenn JO, Siegler EL. Supplementary file: Assessing Electronic Note Quality Using the Physician Documentation Quality Instrument (PDQI-9). 2012 [cited 2013 02/18]; Available from: http://aci.schattauer.de/fileadmin/assets/zeitschriften/ACI/supplementary_material/02_ACI2011110070_Stetson_Appendix.pdf.

77. Merriam-Webster. English-Spanish dictionary. [cited 2013 07/21]; Available from: http://www.spanishcentral.com/.

78. RAE. Diccionario de la Real Academia Española. [cited 2013 07/21]; Available from: http://www.rae.es/rae.html.

79. Lassaque L. Diccionario de falsos cognados inglés – castellano. Buenos Aires2006. Available from: http://www.avlt.com.ar/archivos/libro/29_critic.pdf.

80. Diccionario de la lengua española Espasa-Calpe en wordreference.com. 2005 [cited 2013 07/21].

81. NCI. Interpreting Laboratory Test Results. 2003 [cited 2013 07/21]; Available from: http://www.cancer.gov/cancertopics/factsheet/detection/laboratory-tests.

Appendices

Appendix 1: The PDQI-9 Instrument

Appendix: Physician Documentation Quality Instrument (PDQI-9)

Date: _____ **Author:** _____ **Reviewer:** _____

Note Type (circle): Admit Progress Discharge

Instructions: Please review the chart before assessing the note. Then rate the note on each of the following attributes:

Attribute	Score					Description of Ideal Note
1. Up-to-date	Not at all 1	2	3	4	Extremely 5	The note contains the most recent test results and recommendations.
2. Accurate	Not at all 1	2	3	4	Extremely 5	The note is true. It is free of incorrect information.
3. Thorough	Not at all 1	2	3	4	Extremely 5	The note is complete and documents all of the issues of importance to the patient.
4. Useful	Not at all 1	2	3	4	Extremely 5	The note is extremely relevant, providing valuable information and/or analysis.
5. Organized	Not at all 1	2	3	4	Extremely 5	The note is well-formed and structured in a way that helps the reader understand the patient's clinical course.
6. Comprehensible	Not at all 1	2	3	4	Extremely 5	The note is clear, without ambiguity or sections that are difficult to understand.
7. Succinct	Not at all 1	2	3	4	Extremely 5	The note is brief, to the point, and without redundancy.
8. Synthesized	Not at all 1	2	3	4	Extremely 5	The note reflects the author's understanding of the patient's status and ability to develop a plan of care.
9. Internally Consistent	Not at all 1	2	3	4	Extremely 5	No part of the note ignores or contradicts any other part.
Total Score:						

(Version 1: 11/21/2011)

Figure 13: The PDQI-9 Instrument (76)

Appendix 2: Translation Phase Participants and Members of Expert Committee

Translation Phase Participants	**Qualifications/Title**
FT1 Translator	Bilingual Executive Secretary, academic secretary at Internal Medicine Department at HIBA
FT2 Translator	MD, Internal Medicine and Hypertension specialist at HIBA
BT1 Translator	Master of Science in Environmental Studies, Master in Public Health, research management at OHSU
BT2 Translator	MD, Internal Medicine and Cancer and Blood Disorders specialist at OHSU Hospital
Members of Expert Committee	
Member 1	MD, Internal Medicine and Hypertension specialist at HIBA
Member 2	MD, Internal Medicine and Home care specialist at HIBA
Member 3	MD, Internal Medicine specialist and Pulmonologist at HIBA
Member 4	MD, Internal Medicine and Emergency Medicine specialist at HIBA
Coordinator	MD, Internal Medicine specialist and Medical Informaticist at HIBA, MS candidate at OHSU

FT: Forward translation

BT: Back translation

MD: Medical Doctor

HIBA: Hospital Italiano de Buenos Aires

OHSU: Oregon Health and Science University

Appendix 3: Evaluation of Forward Translations and Creation of the Beta Version Report

Phase: Adaptation	
Objectives: 1.Evaluation of forward translations 2.Creation of beta version	
Date, place: June 28th, 2013, Buenos Aires, Argentina	
People involved: Members of Expert Committee (Members 1, 2, 3 and 4, coordinator)	
Items that did not require discussion: None	
Discrepancies and their resolution	
1.a. Attributes	
Issue	**Resolution**
The word "nota" ("note", in English) is not used in Spanish as a medical word.	The experts chose the word "registro" instead of "nota" or "documento" ("record" instead of "note" or "document"). They agreed that "registro" conveyed the sense of the word "note" in Spanish. "Documento" would be associated with a legal record or the documentation process and not necessarily to the communication process. The experts preferred "registro" to "registro medico" (medical record) because the medical context provides enough clarification.
Some of the attributes in the translations were nouns instead of adjectives.	The experts agreed that attributes should be in the adjective form, if possible.
Harmonization of attribute number 1 (Up-to-date)	The experts considered the word "actualizado" (adjective) to convey in Spanish the sense of the word "up-to-date". They chose the word "actualizado" (adjective) instead of "actualización" (noun).
Harmonization of attribute number 2 (Accurate)	The experts agreed that the word "preciso" (adjective) conveyed in Spanish the sense of the word "accurate". Another translation for "accurate" is "exacto"; however, in this adaptation "preciso" means "cierto" ("true") (77).
Harmonization of attribute number 3 (Thorough)	The experts agreed that the word "completo" (adjective) conveyed in Spanish the sense of the word "thorough". Another translation for "thorough" is "minucioso" ("detailed") (77); however, the description of the ideal note states that "The note is complete and documents all of the issues of importance to the patient." That means the note should be complete enough to be understood by the patient and not fully detailed with irrelevant aspects. This attribute may express the association of two attributes from the original 22-item score:

	complete and important.
Harmonization of attribute number 4 (Useful)	The experts agreed that the word "útil" (adjective) conveyed in Spanish the sense of the word "useful". This attribute may express the association of two attributes from the original 22item score: useful and relevant.
Harmonization of attribute number 5 (Organized)	The experts agreed that the word "organizado" (adjective) communicates in Spanish the sense of the word "organized". This attribute may express the association of two attributes from the original 22-item score: organized and structured.
Harmonization of attribute number 6 (Comprehensible)	The experts agreed that the word "comprensible" (adjective) expresses in Spanish the sense of the word "comprehensible". A synonym of "comprensible" is "inteligible" (77). "Inteligible" is often synonymous with "entendible" (78). The experts preferred "comprensible" to "inteligible" or "entendible". The word "inteligible" sounds too formal and "entendible", too informal. This attribute may express the association of two attributes from the original 22-item score: clear and comprehensible.
Harmonization of attribute number 7 (Succinct)	The experts accepted "conciso" (adjective) as a word that communicates in Spanish the sense of the word "succinct". This attribute may express the association of two attributes from the original 22-item score: concise and brief.
Harmonization of attribute number 8 (Synthesized)	The experts agreed that the word "sintetizado" (adjective) expresses in Spanish the sense of the word "synthesized". A synonym of "sintetizado" is "sintético" (78). The experts preferred "sintetizado" to "sintético" because "sintético" is commonly associated to production of plastic. One of the experts proposed to eliminate this attribute. The final consideration was discussed in the creation of the consolidated version.
Harmonization of attribute number 9 (Internally Consistent)	Both translators translated "Consistent" as "consistente". However, although the word pair "consistent-consistente" is almost identical in form, they have different meanings. They are "false cognates". One correct translation for "consistent" can be "coherente" (79). "Internally" was translated as "internamente", and the experts agreed with the translation. They debated about the order of the words "internamente coherente" vs. "coherente internamente", and did not find any significant differences. Therefore, the experts

	decided to use the words "internamente coherente".
1.b. Score	
Total score	One of the translators translated "score" as "puntuación", and the other translator changed the word order. The experts preferred "puntaje total" to "puntuación total " because that was the word used in the region (78).
1.c. Description of the ideal note	
The word "nota" ("note" in English) is not used in Spanish as a medical word.	The experts chose the word "registro" instead of "nota" or "documento".
"Up-to-date" ideal note	In the description of the ideal note for the attribute "up-to-date", "test results" was translated as "pruebas". Even when that is a correct translation, the experts preferred "estudios" as that is the word that better conveys that concept in Spanish.
"Organized" ideal note	In the description of the ideal note for the attribute "organized", "well-formed" was translated as "bien formada", in both translations. This is correct; however, the experts preferred "bien construido" because it conveys the concept better.
Sentence order, word order	Every description was reviewed and words and sentences were rearranged.
2. Header	
The word "nota" ("note" in English) is not used in Spanish as a medical word.	The experts chose the word "registro" instead of "nota" or "documento".
"Type of note"	The original form used "admit" for "admission note", "progress" for "progress note", and "discharge" for "discharge summary". The translators translated the words literally ("admit"-"admitir", "progress"-"progresar" and "discharge"-"descartar"). Even though the translations were correct, the experts did not approve the translations because the words did not represent what they mean as medical concepts (clinical notes). The experts chose "ingreso", "evolución" and "epicrisis", respectively.
"Chart"	"Chart" was translated as "cuadro" and "cartilla". Even though the translations were correct, the experts did not approve them because the words did not represent what they mean. The experts decided to use "historia clínica" instead.
"Rate the note"	The concept of "rate the note" was translated as "califique (valore) la nota" and "asignar la nota". The experts finally decided to use "califique al

	registro" because it represented better the concept of "rate the note" in Spanish. The discussion was intended to make sure that the point of evaluation was the note and not the entire chart.
3.Title	
Physician Documentation Quality Instrument (PDQI-9)	The translations were confusing because of the word arrangement. Following the suggestion of the coordinator, the experts, after a discussion, agreed to add the word "evaluación" (evaluation) for enhanced understanding. Finally, the title was defined as "Instrumento de Evaluación de la Calidad de la Documentación Médica".

Appendix 4: Forward Translation of Attributes and Description of Ideal Notes, and Creation of the Beta Version

Original English version	Forward Translation 1 (FT1)(Spanish)	Forward Translation 2 (FT2) (Spanish)	Beta version(1st version in Spanish)
Attribute "Description of Ideal Note"	**Atributo** “Descripción de Nota Ideal”	**Atributo** “Descripción de Nota Ideal”	**Atributo** "Descripción del *registro* ideal"
Up-to-date "The note contains the most recent test results and recommendations."	***Actualización*** “La nota contiene los resultados de *pruebas* más recientes y recomendaciones.”	***Actualizado*** “La nota contiene los resultados de las *pruebas* y las recomendaciones más recientes.”	***Actualizado*** "El *registro* contiene los resultados de los *estudios* y recomendaciones más recientes."
Accurate "The note is true. It is free of incorrect information"	***Preciso*** **(certero/exacto)** “La nota es verdadera. Está libre de información incorrecta.”	***Preciso*** “La nota es verdadera. Se encuentra libre de información incorrecta.”	**Preciso** "El *registro* es verdadero. Esta libre de información incorrecta."
Thorough "The note is complete and documents all of the issues of importance to the patient."	***Exhaustivo*** **(minucioso)** “La nota es(tá) completa y documenta todas las cuestiones de importancia del paciente.”	***Completo*** “La nota es completa y documenta todos los temas de importancia del paciente.”	***Completo*** "El *registro* es completo y documenta todos los temas de importancia acerca del paciente."
Useful "The note is extremely relevant, providing valuable information and/or analysis."	**Util** “La nota es extremadamente relevante, proporcionando valiosa información y/o análisis.”	**Útil** “La nota es extremadamente relevante, proveyendo valiosa información y/o análisis.”	**Útil** "El *registro* es extremadamente relevante, proporcionado información y/o análisis valiosos."
Organized "The note is well-formed and structured in a way that helps the reader understand the patient˝s clinical course."	**Organizado** “La nota está bien formada y estructurada de forma tal que ayuda al lector a comprender el curso clínico del paciente.”	**Organizado** “La nota está bien formada y estructurada en una forma que ayuda al lector entender el curso clínico del paciente.”	**Organizado** "El *registro* está bien *construido* y estructurado de forma tal que ayuda al lector entender el curso clínico del paciente."
Comprehensible "The note is clear, without ambiguity or sections that are difficult to understand."	***Comprensible*** “La nota es clara, sin ambigüedad ni secciones que son difíciles de entender (comprender).”	***Entendible*** “La nota es clara, sin ambigüedades o secciones difíciles de entender.”	**Comprensible** "El *registro* es claro, sin ambigüedad ni secciones difíciles de entender."
Succinct "The note is brief, to the point, and without redundancy.”	***Sucinto*** “La nota es breve, va al punto, y sin redundancia.”	***Conciso*** “La nota es breve, al punto y sin redundancias.”	***Conciso*** "El *registro* es breve, no redundante y va al punto."
Synthesized "The note reflects the author˝s understanding of the patient˝s status and ability to develop a plan of care."	***Sintetizado*** “La nota refleja la comprensión del autor del estado del paciente y habilidad de desarrollar un plan de atención.”	***Sintético*** “La nota refleja la idea que tiene el autor sobre el estado del paciente y la habilidad para desarrollar un plan de cuidado.”	***Sintetizado*** "El *registro* refleja la comprensión del autor sobre el estado del paciente y su habilidad para desarrollar un plan de cuidado."
Internally Consistent "No part of the note ignores or contradicts any other part.”	**Internamente *consistente*** “No existe parte de la nota que ignore ni contradiga alguna otra parte de la misma.”	**Internamente *consistente*** “No hay una parte de la nota que ignore o contradiga otra parte.”	**Internamente *coherente*** "Ninguna parte del *registro* ignora o contradice a otra parte."
Total *Score*	*Puntuación* total	*Score* Total	*Puntaje* Total

Appendix 5: Header and Title Forward Translation and Creation of the Beta Version

Original English version	**Forward Translation 1 (FT1)(Spanish)**	**Forward Translation 2 (FT2) (Spanish)**	**Beta version(1st version in Spanish)**
Header			
Date: _____ Author: _____ Reviewer: _____	Fecha: Autor: Revisor:	Fecha: _____ Autor: _____ Revisor: _____	Fecha: _____ Autor: _____ Revisor: _____
Note Type (***circle***): ***Admit*** ***Progress*** ***Discharge***	***Nota*** Tipo (círculo): ***Admitir (aceptar)*** ***Progreso (Evaluación Etapa)*** ***Descargo (desempeño)***	Tipo de ***Nota*** (***círculo***): ***Admitir*** ***Progresar*** ***Descartar***	Tipo de ***Registro*** (***marque con un círculo***): ***Ingreso*** ***Evolución*** ***Epicrisis***
Instructions: Please review the ***chart*** before assessing the ***note***. Then ***rate the note*** on each of the following attributes:	Instrucciones: Por favor, revise (revea) el ***cuadro*** antes de evaluar la ***nota***. Luego, ***califique (valore) la nota*** sobre cada uno de los siguientes atributos:	Instrucciones: Por favor revisar la ***cartilla*** antes de evaluar la ***nota***. Luego ***asignar la nota*** según cada uno de los siguientes atributos:	Instrucciones: Por favor revise la ***Historia Clínica*** antes de evaluar el ***registro***. Luego ***califique al registro*** según cada uno de los siguientes atributos:
Title			
Physician Documentation Quality Instrument (PDQI9)	Documentación de Calidad de Instrumento del Médico	Instrumento de Documentación de Calidad Médica (IDCM-9)	Instrumento de *Evaluación* de la Calidad de la Documentación Médica

Appendix 6: Evaluation of Backward Translations and Creation of the Consolidated Version Report

Phase: Adaptation	
Objectives: 1.Evaluation of backward translations 2.Creation of the consolidated version	
Date, place: July 4th 2013, Buenos Aires, Argentina	
People involved: Members of Expert Committee (Members 1, 2, 3 and 4, coordinator)	
Member 4 was absent. The coordinator and member 4 met by phone and reviewed the topics before the meeting with the rest of the experts. The conclusions of the meeting with the experts were communicated to Member 4.	
Items that did not require discussion: None	
Discrepancies and their resolution	
1.a. Attributes	
Issue	**Resolution**
Harmonization of attribute number 1 (Actualizado)	The word "actualizado" (adjective) was translated as "current" and "up to date". Both words convey in English the sense of the word "actualizado". The word in the original version was „up to date" and one of the translators used this word. Finally, the expert maintained the word "actualizado" in the consolidated version. All of the experts agreed that this attribute is relevant and has to be in the consolidated version.
Harmonization of attribute number 2 (Preciso)	The word "preciso" (adjective) was translated as "precise" and "accurate". Both words convey in English the sense of the word "preciso" and are synonyms (77). The word in the original version was "accurate" and one of the translators used this word. The expert continued to use the word "preciso" in the consolidated version. All of the experts agreed that this attribute is relevant and has to be in the consolidated version.
Harmonization of attribute number 3 (Completo)	"Completo" (adjective) was translated as "complete". The word conveys in English the sense of the word "completo" (77). The original word was "thorough"; however, none of the translators used "thorough". The consolidated version maintained the word "completo" because the description of the ideal record states that the record should be complete in terms of relevant information about the patient and not fully detailed with irrelevant aspects. All of the experts agreed that this attribute

	is relevant and has to be in the consolidated version.
Harmonization of attribute number 4 (Útil)	"Útil" (adjective) was translated as „useful" and was considered to be correctly translated. The consolidated version maintained the word "útil". All of the experts agreed that this attribute is relevant and has to be in the consolidated version.
Harmonization of attribute number 5 (Organizado)	The translations of "organizado" (adjective) were "organized" (adjective) and "organization" (noun). The translation was considered correct in the first case and acceptable in the second one, because both words are related. The original version of the word was „organized" and one of the translators used this word. The consolidated version maintained the word "organizado". All of the experts agreed that this attribute is relevant and has to be in the consolidated version.
Harmonization of attribute number 6 (Comprensible)	The translations of "comprensible" (adjective) were "understandable" (adjective) and "clarity" (noun). The word in the original version is "comprehensible". One of the translators used a synonym, while the other one did not. This can be explained by the fact that "comprehensible" represents two attributes: 1) "clear" and 2) "understandable". This is shown in the description of the ideal note in the original version, "The note is clear, without ambiguity or sections that are difficult to understand". The experts deliberated about this point and decided that a record should be clear so that it can be understood, and that makes a record "comprehensible". Finally, the consolidated version maintained the word "comprensible". All of the experts agreed that this attribute is relevant and has to be in the consolidated version. Note: Translator 2 chose "clarity" because he believed that it was easier and simple to understand. He agreed on the use of "clear" instead.
Harmonization of attribute number 7 (Conciso)	The translations of "conciso" (adjective) were "concise" (adjective) and "brevity" (noun). The definition of "concise" is something "marked by brevity of expression or statement" (77). The word in the original version is "succinct". Following the word from the beta version, one of the translations was literal, but the second one was not.

	One of the experts argued that “succinct” could represent two attributes: 1) “brief” and 2) “concise” (“breve, conciso y preciso”) (80). After a debate, the experts considered carefully how to represent this concept. Finally, they agreed to use the original word “succinct” translated. The consolidated version opted for the word “sucinto” instead. All of the experts agreed that this attribute is relevant and has to be in the consolidated version. Note: Translator 2 chose “brevity” because he believed it was easier and simple to understand. He agreed on the use of “brief” instead.
Harmonization of attribute number 8 (Sintetizado)	The translations of “sintetizado” (adjective) were “synthesized” (adjective) and “well-reasoned” (adverb + adjective). The translation was considered correct in the first case. The second one was considered an interpretation in addition to a simple translation. In the original version the word was “synthesized” and one of the translators used this word. Translator 2 clarified the reason why he used the concept “well-reasoned”. He considered that “synthesized” did not represent the concept correctly. The relevance of this attribute was considered too. Three out of four experts agreed that the attribute should be in the consolidated version. The argument for avoiding its use was to have a shorter version. The argument for using it was that this is an essential attribute that shows that the clinician understands what happened with the patient, makes a plan of care and can communicate it. After a debate, the experts analyzed carefully how to represent this concept, especially taking into account the meaning of “well-reasoned”. According to the original description of the ideal note (“The note reflects the author‟s understanding of the patient‟s status and ability to develop a plan of care.") and the original word (“synthesized”), this attribute represents the physician‟s skill to make a synthesis of the patient‟s status based on which a plan of care is to be developed. One of the experts suggested “interpretado”. This concept is known locally as “interpretación” and its translation to English is “interpretation”. The closest meaning is related to “interpreting laboratory test results”,

	where the information of laboratory tests helps doctors make decisions (81). “Well-reasoned”, “reasoned” or “synthesized” do not communicate the concept in its entirety. The experts deliberated about this point and decided that “interpretado” should be in the consolidated version.
Harmonization of attribute number 9 (Internamente coherente)	The translations of “internamente coherente” were “internally consistent” and “consistent”. The words in the original version were „‟internally consistent‟‟ and one of the translators used these words. The consolidated version included the concept of “internamente coherente”. All of the experts agreed that this attribute is relevant and has to be in the consolidated version.
Order of attributes	Member 4 suggested evaluating in the future the order of the attributes to come up with a simplified way of using them.
1.b.Score	
“Puntaje total”	“Puntaje total” was translated as “total score”. The experts agreed with the translation. The consolidated version included the concept “puntaje total”.
1.c. Description of the ideal note	
"Descripción del registro ideal"	In “descripción del registro ideal”, “registro” was translated as “record” and “medical record entry”. One of the translators made a point: “If the reviewer was looking at the entire patient medical record, I would use „‟chart‟‟ for „‟registro‟‟. If he was just looking at a small section of the chart, I would use „‟note‟‟. If it could be either, I would use „‟record‟‟ (more generic).”
“Actualizado” ideal note	The experts agreed about the back translations. The consolidated version emphasized that the record should contain the status of the patient: "El registro contiene los resultados de los estudios, el estado clínico actual del paciente y recomendaciones más recientes."
“Útil” ideal note	The experts agreed about the back translations. The consolidated version emphasized that usefulness is for the sake of patient care. "El registro es extremadamente relevante y útil, proporcionando información y/o análisis valiosos para el cuidado del paciente."

“Organizado” ideal note	The back translations were correct. After the review of the beta version, the experts agreed to change the word “contruido” to “estructurado”: "El registro está estructurado de forma tal que ayuda al lector a entender el curso clínico del paciente."
“Sucinto” ideal note	The back translations were correct. However, the attribute was changed. The consolidated version added the word “concise” (concise) in the description to reinforce the concept. “El registro es breve y conciso, no redundante y va al punto.”
2. Header	
The word “nota” (“note” in English) is not used in Spanish as a medical word.	The experts chose the word “registro” instead of “nota” or “documento”.
“Tipo de registro”	“Tipo de registro” was back translated as “type of record” and “type of entry”. The original words were “type of note”. The experts accepted the translations from the beta version. There were discrepancies in the translations of the type of record. One of the translators expressed his doubts and suggested some options: “admission” for “ingreso”, “evaluation”, “recovery” or “progress note” for “evolución”, and “discharge summary” for “epicrisis”. Experts debated about this point and concluded that discrepancies could be fewer if the words used in the beta version were more generic: for instance, “registro al ingreso”, “registro de evolución diaria”, and “registro al alta”, instead of “ingreso”, “evolución”, and “epicrisis”.
“Historia clínica”	“Historia clínica” was translated as “clinical history”. One of the translators clarified and expanded her translation: “the record/chart is the entire thing; the clinical history is a section of the chart (...)” The translation should depend on the context. A generic word for “historia clínica” is “registro medico” (medical record) but such term is not commonly used.
3. Title	
Instrumento de Evaluación de la Calidad de la Documentación Médica	“Physician Documentation Quality Instrument (PDQI-9)” was the original title of this score. The beta version was “Instrumento de Evaluación de la Calidad de la Documentación Médica”. This

	version was translated as “Instrument (tool) for Evaluating Medical Documentation Quality” and “Instrument to Evaluate the Quality of Medical Documentation”. Experts agreed that it expressed the meaning of the beta version and it was the same for the consolidated version. There was consensus about maintaining the use of word “documentation” and the final title of the instrument will be defined with its use.

Appendix 7: Backward Translations of Attributes and Description of Ideal notes, and Creation of the Consolidated Version

Beta version(1st version in Spanish)	Back-translation 1(BT1)	Back-translation 2 (BT2)	Consolidated version(2nd version in Spanish)
Atributo "Descripción del *registro* ideal"	**Attribute** "Description of the ideal *record*"	**Attribute** "Description of the ideal *medical record entry*"	**Atributo** "Descripción del *registro* ideal"
Actualizado "El *registro* contiene los resultados de los estudios y recomendaciones más recientes."	**Current** "The *record* contains the most recent *test* results and recommendations."	**Current/Up to date** "The *medical record* contains the most recent study results and recommendations."	**Actualizado** "El *registro* contiene los resultados de los estudios, *el estado clínico actual del paciente* y recomendaciones más recientes."
Preciso "El *registro* es *verdadero*. Esta libre de información incorrecta."	**Precise** "The *record* is *accurate* and free of incorrect information"	**Accurate** "The *medical record* is *accurate*. It is free of incorrect information"	**Preciso** "El *registro* es verdadero. Esta libre de información incorrecta"
Completo "El *registro* es completo y documenta todos los temas de importancia acerca del paciente."	***Complete*** "The *record* is complete and documents all important issues regarding the patient."	**Complete** "The *medical record* is complete and documents all the important issues regarding the patient."	**Completo** "El *registro* es completo y documenta todos los temas de importancia acerca del paciente."
Útil "El *registro* es extremadamente relevante, proporcionado información y/o análisis valiosos."	**Useful** "The *record* is extremely *relevant*, with suitable information and/or valid analysis."	**Useful** "The *medical record* is useful, providing valuable information and analysis."	**Útil** "El *registro* es extremadamente relevante y útil, proporcionado información y/o análisis valiosos *para el cuidado del paciente.*"
Organizado "El *registro* está bien *construido* y estructurado de forma tal que ayuda al lector entender el *curso* clínico del paciente."	**Organized** "The *record* is *well-constructed* and structured in a way that helps the reader understand the clinical course of the patient."	***Organization*** "The *medical record* is well *organized* and structure in a manner that helps the reader understand the patients" clinical *trajectory*"	**Organizado** "El registro está *estructurado* de forma tal que ayuda al lector a entender el curso clínico del paciente."
Comprensible "El *registro* es claro, sin ambigüedad ni secciones difíciles de entender."	***Understandable*** "The *record* is clear, without ambiguity or sections that are difficult to understand."	***Clarity*** "The *medical record* is clear, without ambiguity nor entries that are difficult to understand."	**Comprensible** "El *registro* es claro, sin ambigüedad ni *partes* difíciles de entender"
Conciso "El *registro* es *breve*, no redundante y va al punto."	***Concise*** "The *record* is *brief*, to the point, and without redundancies."	***Brevity*** "The *medical record* is *concise*, not redundant y gets to the point."	***Sucinto*** "El *registro* es breve *y conciso*, no redundante y va al punto."
Sintetizado "El *registro* refleja la comprensión del autor sobre el estado del paciente y su habilidad para desarrollar un plan de cuidado."	***Synthesized*** "The *record* reflects the author"s understanding of the *patient's condition* and ability to develop a plan of care."	***Well Reasoned*** "The *medical record* reflects the author"s understanding of the patient"s condition and his/her ability to develop a plan of care."	***Interpretado*** "El *registro* refleja la comprensión del autor sobre el estado del paciente y su habilidad para desarrollar un plan de cuidado."

Internamente *coherente* "Ninguna parte del *registro* ignora o contradice a otra parte."	***Internally Consistent*** "No part of the *record* ignores or contradicts another part."	***Consistent*** "No part of the *medical record* ignores or contradicts another part"	**Internamente coherente** "Ninguna parte del *registro* ignora o contradice a otra parte."
Puntaje Total	Total Score	Total Score	Puntaje Total

Appendix 8: Header and Title Backward Translations and Creation of the Consolidated Version

Beta version(1st version in Spanish)	**Back-translation 1 (BT1)**	**Back-translation 2 (BT2)**	**Consolidated version(2nd version in Spanish)**
Header			
Fecha: _____ Autor: _____ Revisor: _____	Date: _____ Author: _____ Reviewer: _____	Date: _____ Author: _____ Reviewer: _____	Fecha: _____ Autor: _____ Revisor: _____
Tipo de ***Registro*** (***marque con un círculo***): ***Ingreso*** ***Evolución*** ***Epicrisis***	Type of ***Record (please circle):*** ***Admission*** ***Evaluation (Recovery? Progress Note?)*** ***Discharge Summary***	Type of ***Entry (Mark with a circle):*** ***Initial*** Evolución (¿?) Epicrisis (¿?)	Tipo de ***Registro*** (***marque con un círculo***): ***Ingreso*** ***Evolución*** ***Epicrisis***
Instrucciones: Por favor revise la ***Historia Clínica*** antes de evaluar el ***registro***. Luego ***califique al registro*** según cada uno de los siguientes atributos:	Instructions: Please review the ***clinical history*** prior to evaluating the ***medical record entry***. ***Grade the record*** according to each of the following attributes:	Instructions: Please review the ***Clinical History*** before evaluating the ***medical record entry***. Then, ***grade the record*** according to each of the following attributes:	Instrucciones: Por favor revise la ***Historia Clínica*** antes de evaluar el ***registro***. Luego ***califique al registro*** según cada uno de los siguientes atributos:
Title			
Instrumento de Evaluación de la Calidad de la Documentación Médica	Instrument (tool) for Evaluating Medical Documentation Quality	Instrument to Evaluate the Quality of Medical Documentation	Instrumento de Evaluación de la Calidad de la Documentación Médica

Appendix 9: Comparison of the Original English Version with the Backward Translation

Original English version	Back-translation 1(BT1)	Back-translation 2 (BT2)
Attribute "Description of Ideal *Note*"	**Attribute** "Description of the ideal *record*"	**Attribute** "Description of the ideal *medical record entry*"
Up-to-date "The *note* contains the most recent *test* results and recommendations."	**Current** "The *record* contains the most recent *test* results and recommendations."	**Current/Up to date** "The *medical record* contains the most recent *study* results and recommendations."
Accurate "The *note* is *true*. It is free of incorrect information"	**Precise** "The *record* is *accurate* and free of incorrect information"	**Accurate** "The *medical record* is *accurate*. It is free of incorrect information"
Thorough "The *note* is complete and documents all of the issues of importance to the patient."	***Complete*** "The *record* is complete and documents all important issues regarding the patient."	***Complete*** "The *medical record* is complete and documents all the important issues regarding the patient."
Useful "The *note* is extremely *relevant*, providing valuable information and/or analysis."	**Useful** "The *record* is extremely *relevant*, with suitable information and/or valid analysis."	**Useful** "The *medical record* is *useful*, providing valuable information and analysis."
Organized "The *note* is *well-formed* and structured in a way that helps the reader understand the patient‟s clinical course."	**Organized** "The *record* is *well-constructed* and structured in a way that helps the reader understand the clinical course of the patient."	***Organization*** "The *medical record* is *well organized* and structure in a manner that helps the reader understand the patients‟ clinical trajectory"
Comprehensible "The *note* is clear, without ambiguity or sections that are difficult to understand."	***Understandable*** "The *record* is clear, without ambiguity or sections that are difficult to understand."	***Clarity*** "The *medical record* is clear, without ambiguity nor entries that are difficult to understand."
Succinct "The *note* is *brief*, to the point, and without redundancy."	***Concise*** "The *record* is *brief*, to the point, and without redundancies."	***Brevity*** "The *medical record* is *concise*, not redundant y gets to the point."
Synthesized "The *note* reflects the author‟s understanding of the *patient's status* and ability to develop a plan of care."	***Synthesized*** "The *record* reflects the author‟s understanding of the *patient's condition* and ability to develop a plan of care."	***Well Reasoned*** "The *medical record* reflects the authors understanding of the *patient's condition* and his/her ability to develop a plan of care."
Internally Consistent "No part of the *note* ignores or contradicts any other part."	***Internally Consistent*** "No part of the *record* ignores or contradicts another part."	***Consistent*** "No part of the *medical record* ignores or contradicts another part"
Total Score	Total Score	

Appendix 10: Comparison of the Original English Version with the Beta and Consolidated Versions

Original English version	Beta version(1st version in Spanish)	Consolidated version(2nd version in Spanish)
Attribute "Description of Ideal *Note*"	**Atributo** "Descripción del *registro* ideal"	**Atributo** "Descripción del *registro* ideal"
Up-to-date "The *note* contains the most recent test results and recommendations."	**Actualizado** "El *registro* contiene los resultados de los estudios y recomendaciones más recientes."	**Actualizado** "El *registro* contiene los resultados de los estudios, *el estado clínico actual del paciente* y recomendaciones más recientes."
Accurate "The *note* is true. It is free of incorrect information"	**Preciso** "El *registro* es verdadero. Esta libre de información incorrecta."	**Preciso** “El *registro* es verdadero. Esta libre de información incorrecta”
Thorough "The *note* is complete and documents all of the issues of importance to the patient."	**Completo** "El *registro* es completo y documenta todos los temas de importancia acerca del paciente."	**Completo** “El *registro* es completo y documenta todos los temas de importancia acerca del paciente.”
Useful "The *note* is extremely relevant, providing valuable information and/or analysis."	**Útil** "El *registro* es extremadamente relevante, proporcionado información y/o análisis valiosos."	**Útil** "El *registro* es extremadamente relevante y útil, proporcionado información y/o análisis valiosos *para el cuidado del paciente*."
Organized "The *note* is *well-formed* and *structured* in a way that helps the reader understand the patient‟s clinical course."	**Organizado** "El registro está *bien construido* y *estructurado* de forma tal que ayuda al lector entender el curso clínico del paciente."	**Organizado** "El registro está *estructurado* de forma tal que ayuda al lector a entender el curso clínico del paciente."
Comprehensible "The *note* is clear, without ambiguity or sections that are difficult to understand."	**Comprensible** "El *registro* es claro, sin ambigüedad ni *secciones* difíciles de entender."	**Comprensible** "El *registro* es claro, sin ambigüedad ni *partes* difíciles de entender"
Succinct "The *note* is brief, to the point, and without redundancy.”	***Conciso*** "El *registro* es breve, no redundante y va al punto."	***Sucinto*** “El *registro* es breve *y conciso*, no redundante y va al punto.”
Synthesized "The *note* reflects the author‟s understanding of the patient‟s status and ability to develop a plan of care."	***Sintetizado*** "El *registro* refleja la comprensión del autor sobre el estado del paciente y su habilidad para desarrollar un plan de cuidado."	***Interpretado*** "El *registro* refleja la comprensión del autor sobre el estado del paciente y su habilidad para desarrollar un plan de cuidado."
Internally Consistent "No part of the *note* ignores or contradicts any other part.”	**Internamente coherente** "Ninguna parte del *registro* ignora o contradice a otra parte."	**Internamente coherente** “Ninguna parte del *registro* ignora o contradice a otra parte.”
Total Score	Puntaje Total	Puntaje Total

Appendix 11: Matrix of Agreement Among the Different Versions of the PDQI-9

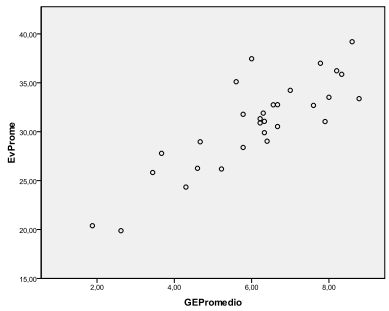
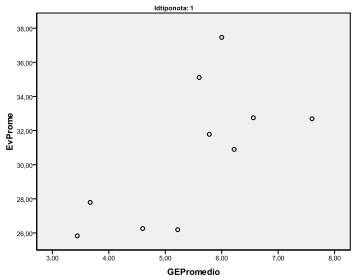
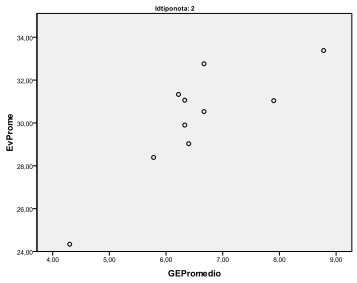
	Or/ft1	Or/ft2	ft1/ft2	Or/B	B/bt1	B/bt2	bt1/bt2	Or/bt1	Or/bt2	B/C	Or/C	A + PA (%)
Review	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	
Title	NA	NA	NA	NA	A	A	A	NA	NA	A	NA	42,86
Header	NA	NA	NA	A	PA	NA	NA	PA	NA	A	A	57,14
Instructions	NA	NA	NA	A	NA	NA	PA	PA	PA	A	A	64,29
Attribute 1	NA	A	NA	A	A	A	PA	PA	A	A	A	85,71
Attribute 2	A	A	PA	PA	A	PA	NA	PA	A	A	PA	92,86
Attribute 3	NA	NA	PA	PA	A	A	A	NA	NA	A	PA	71,43
Attribute 4	A	A	A	A	A	A	A	A	A	A	A	100,00
Attribute 5	A	A	A	A	A	NA	NA	A	NA	A	A	78,57
Attribute 6	A	A	A	A	PA	NA	NA	PA	NA	A	A	78,57
Attribute 7	A	A	PA	PA	A	NA	NA	PA	NA	PA	A	78,57
Attribute 8	A	NA	PA	A	A	NA	NA	A	NA	NA	NA	57,14
Attribute 9	NA	NA	A	A	A	PA	NA	A	PA	A	A	78,57
Desc.1	A	A	A	PA	A	A	A	A	A	PA	PA	100,00
Desc.2	A	A	A	A	PA	A	A	PA	A	A	A	100,00
Desc.3	A	A	A	A	A	A	A	A	A	A	A	100,00
Desc.4	A	A	A	A	A	A	A	A	A	PA	PA	100,00
Desc.5	A	A	A	PA	A	A	A	PA	A	A	A	100,00
Desc.6	A	A	A	A	A	A	A	A	A	A	A	100,00
Desc.7	A	A	A	A	A	A	A	A	A	A	A	100,00
Desc.8	A	A	A	A	A	A	A	A	A	A	A	100,00
Desc.9	A	A	A	A	A	A	A	A	A	A	A	100,00
Total Score	A	A	A	A	A	A	A	A	A	A	A	100,00
A	16	16	14	16	18	14	13	12	13	18	16	
PA	0	0	4	5	3	2	2	8	2	3	4	
NA	6	6	4	1	1	6	7	2	7	1	2	
A + PA	72,73%	72,73%	81,82%	95,45%	95,45%	72,73%	68,18%	90,91%	68,18%	95,45%	90,91%	

Or=original version; ft1= forward translation 1; ft2= forward translation 2; B= beta version; bt1= backward translation1; bt2= backward translation2; C= consolidated version

A= agreement; PA= partial agreement

Desc. = Description

Appendix 12: Criterion Validity (Comparison with GSs)

Item	Pearson Correlation coefficient	p value	
Global (30)	0.85	<0.001	
Admission notes (10)	0.97	<0.001	
Progress notes (10)	0.84	0.003	
Discharge summaries (10)	0.67	0.035