A holistic-componential model for assessing translation student performance and competency*

Malcolm Williams
Universidad de Ottawa
malcolmwilliams@rogers.com

Abstract:
Translation quality assessment (TQA) tools frequently come under attack because of the myriad variables involved in TQA: the definition, number and seriousness of errors, the purpose of the assessment, evaluator competence and reliability, the client's or end user's requirements, deadlines, complexity of the TQA model, etc. In recent years, progress in factoring in these variables and achieving greater reliability and validity has been achieved through functionalist, criterion-referenced models proposed by Colina (2008, 2009) and others for the assessment of professional translation quality, even though they have come under attack from proponents of the normative assessment model (Anckaert et al., 2008, 2009). At the same time, progress has been made in student assessment through the holistic, criterion-referenced approaches developed by education theorists Wiggins (1998) and Biggs and Tang (2007) ─ approaches that have been applied to translation by Kelly (2005). In this article, the author proposes a "holistic-componential" model for translation student assessment. Based on a combination of Colina's functionalist translation assessment model and the holistic student assessment model and drawing on definitions of professional standards applied in North America, it is designed to rectify some of the perceived shortcomings of the conventional quantitative, error-based marking schemes, those of the more "impressionistic" schemes, and even those of criterion-referenced models.
Key Words: criterion-referenced assessment, holistic-componential model, functionalist translation assessment model, holistic student assessment model

Résumé :
Les divers outils d'appréciation de la qualité des traductions font souvent l'objet de critiques en raison de la multiplicité de variables en cause, dont la définition, le nombre et la gravité des fautes, le but de l'évaluation, la compétence et la fiabilité de l'évaluateur, les exigences du client ou de l'utilisateur final, les délais et la complexité du modèle d'évaluation. Or, au cours des dernières années, le modèle fonctionnaliste d'appréciation critériée de la qualité traductionnelle proposé par Colina (2008, 2009), entre autres, a permis de tenir compte de plusieurs variables et d'offrir donc une fiabilité et une validité accrues, même s'il a encouru les foudres des tenants de l'évaluation normative en traduction (Anckaert et coll., 2008, 2009). Parallèlement, des progrès ont été réalisés en évaluation éducative grâce aux approches holistiques et critérielles élaborées par Wiggins (1998) et par Biggs et Tang (2007), chercheurs dans le domaine des sciences de l'éducation. Dans le présent article, l'auteur propose l'intégration du modèle fonctionnaliste à la démarche holistique en s'inspirant des définitions de normes professionnelles appliquées en Amérique du Nord, et ce, en vue de créer une modèle « holistique et componentiel » d'évaluation de la performance et des compétences des apprentis traducteurs. Un tel modèle aurait pour objet de corriger certaines lacunes qu'on reproche aux grilles de correction quantitatives traditionnelles, aux schémas de correction dits impressionnistes, ainsi qu'aux barèmes critériés.
Mots clé : evaluation critériée, modèle holistique et componentiel, modèle d'évaluation, fonctionnaliste, démarche holistique

*This paper makes part of the research Professor Willams carries out concerning the evaluation of translations within the Translation Programme at Université d'Ottawa.

Resumen:
Las diversas herramientas para evaluar la calidad de las traducciones se critican a menudo, debido a las múltiples variables que implica, entre ellas la definición, el número y la gravedad de las faltas, el propósito de la evaluación, la competencia y la fiabilidad del evaluador, las exigencias del cliente o del usuario final, el tiempo y la complejidad del modelo de evaluación. No obstante, en los últimos años, el modelo funcionalista de la evaluación de la calidad de la traducción, basado en criterios establecidos, propuesto por Colina (2008, 2009), entre otros modelos, ha permitido tener en cuenta diferentes variables y, por lo tanto, proporcionar una mayor fiabilidad y validez incluso si es criticado por de los partidarios de evaluación normativa en la traducción (Anckaert et al., 2008, 2009). Asimismo, se han logrado avances en la evaluación educativa por medio de un enfoque holístico y basado en criterios establecidos, desarrollado por Wiggins (1998) y Biggs y Tang (2007), investigadores en el campo de la educación científica. En este artículo, el autor propone la integración del modelo funcionalista con el enfoque holístico, basado en la definición de las normas profesionales aplicables en América del Norte, con el fin de crear un modelo de evaluación "integral y componencial" del desempeño y las habilidades de los traductores en formación. Este modelo corregiría las deficiencias atribuidas a los tradicionales esquemas de corrección cuantitativos, a los esquemas de corrección "impresionistas" y a los modelos basados en criterios establecidos.
Palabras claves: evaluación basada en criterios establecidos, modelo holístico y componencial, modelo de evaluación funcionalista, enfoque holístico

1. Introduction

Whereas there is general agreement about the need for a translation to be "good," "satisfactory" or "acceptable," the definition of acceptability and of the means of determining it are matters of ongoing debate. This state of affairs is attributable to the many factors and variables involved in developing and implementing a TQA model — factors and variables that prompt professionals and academics alike to question the model's validity and reliability.

In recent years, Colina and others have proposed models for professional translation that can be defined as functionalist in that they explicitly factor in the function of the translation and the user's needs and expectations. Indeed, the dynamic quality framework recently developed by the Translation Automation User Society (TAUS) draws heavily on the functionalist approach, applying parameters such as utility, time, sentiment (importance of and impact on brand image), readability, adequacy and fluency (TAUS, 2011). In that sense, the models represent an advance in TQA research. As in the case of a number of earlier models (for example, Larose, 1998; Williams, 2004), the functionalist model requires assessment against several quality criteria as a means of taking full account of function and need and judging the usability of the translation for the target reader. That is why Colina (2008; 2009) refers to her proposal as a "componential-functionalist approach." My goal in this article is to build on this approach by drawing on Wiggins' theory of educative assessment, (1998) Biggs and Tang's constructive alignment perspective (2007), and the definitions of translation quality standards applied by North American professional translators' associations in order to develop a "holistic-componential" model for the assessment of translation student performance and competency — one that would enhance the validity and reliability of the assessor's judgments.

2. Validity and Reliability

Naturally, we strive to be as objective as possible in designing and applying TQA models, and to be successful, we must ensure that our TQA models and procedures pass the test of **validity** and **reliability**.

Validity is the extent to which an evaluation measures what it is designed to measure, such as translation skills (construct validity). Content validity is the extent to which an evaluation covers the skills necessary for performance. For example, is the content of a translation examination an appropriate sample of the content of the course? Does the result of the evaluation accurately predict future performance (predictive validity)?

Reliability is the extent to which an evaluation produces the same results when administered repeatedly to the same population under the same conditions. Thus a TQA system is reliable if evaluators' decisions are consistent and criteria are stable. Are there biases or undue variations in results over time? Is there a mechanism for ensuring that evaluators do not fluctuate between excessive rigour (purism) and extreme flexibility (laxness)? Is the evaluator always objective? Are quality requirements clearly enough defined for decisions on borderline cases to be made with consistency and ample justification?

The *validity* of quantitative TQA schemes, which are used to rate a translation according to the number and seriousness of errors detected, has been challenged because they tend to ignore the macrotextual features of the target text and the fact that a translation with more errors than another may nonetheless may of better overall quality and meet the client's requirements more effectively. A satisfactory model must therefore go beyond quantification. At the same time, the validity of criterion-referenced models such as those of Nord (1991) and House (1997) has been called into question because of the difficulty of moving from an assessment against *each* parameter to an *overall* quality rating for the translation. This problem, along with the avoidance of any quantitative assessment, has opened up the resulting criterion-referenced models to charges of selecting criteria and target competencies subjectively and of engaging in a "holistic (intuitive-impressionistic" method of evaluation (Eyckmans, Anckaert and Segers, 2009, p. 73).

Critics question the *reliability* of TQA schemes for a number of reasons:

- The evaluator may not possess the translation/subject-field experience or evaluation competencies required.
- Evaluators often apply an inappropriate level of target language rigour.
- Evaluators are often inconsistent in determining what is a major or critical error as opposed to a minor error.

- The teacher/assessor knows the students, so the halo effect may jeopardize the objectivity of the marking process.
- Variations in the administrative settings of examinations, tests and other conditions of performance have not been minimized to an acceptable level.

3. Educative assessment and constructive alignment as sources of student performance and competency assessment

Education assessment schemes have been the target of many of the same charges levelled at TQA schemes. Education theorists, including Wiggins and Biggs and Tang, have highlighted an over-reliance on quantitative measurement and on norm-referenced evaluation: students are rated against one another rather than against an achievement target; a particular percentage mark reflects a student's performance relative to his or her peers, not relative to a learning objective; and the criteria against which the student is to be assessed are vague, if they exist at all. What I propose to do is to explore how Wiggins's "educative assessment" approach and Biggs and Tang's "constructive alignment" approach can be combined with the componential-functionalist approach (CFA) to produce a productive educational TQA model. In particular, I will build on CFA as the basis for an educational assessment "dashboard" by exploiting the concepts of *achievement target*, *intended learning outcome*, *standard*, *criterion*, *indicator* and *rubric* to establish the level of quality that needs to be demonstrated for a given grade to be assigned. In this endeavour I will be echoing ideas formulated by Kelly (2005) and Angelelli (2009), who have applied some of the above concepts to translator training and testing.

Wiggins bases his approach on a "logic of assessment design" that he deliberately calls "backward." Challenging a conventional instructional procedure in which teachers develop course activities and then come up with testing and evaluation tools in light of those activities, he proposes a sequence in which assessment design precedes the preparation of teaching and learning activities.

In principle, application of educative assessment should yield not only an evaluation of student performance in relation to targets and standards but also information on areas for improvement that will help students progress toward their achievement targets.

An educative assessment system, Wiggins contends, is "built on a bedrock of meaningful performance tasks that are credible and realistic (authentic), hence engaging to students" (1998, 12). By "authentic," Wiggins means tasks (assignments and tests) that replicate what professionals do "in the real world" and the problems and performance challenges they face, require real-world use of knowledge, and enable students to "do" the subject concerned. In that sense, the translation assignment or test, if carefully constructed and carried out in appropriate conditions, can reflect very closely what professionals do. Furthermore, the translation teacher/assessor can enhance authenticity by ensuring that usability of the translation in a "real world"

context is the overriding yardstick. This is where a carefully constructed set of quality standards and criteria directly related to professional standards comes into play.

3.1 Achievement target or intended learning outcome

The terminology has varied since Bloom's seminal work on the taxonomy of educational objectives (Bloom, 1956). For example, Biggs and Tang (2007, p. 5) refer to outcomes-based education and recommend the establishment of "intended learning outcomes" as the results that teachers and students should aim for. According to the constructive alignment theory that they espouse, learners use their own activity to construct their knowledge or other outcome by actually "doing" and learning the subject concerned. "The 'alignment' in constructive alignment reflects the fact that the learning activity in the intended outcomes, expressed as a verb, needs to be activated in the teaching of the outcome to be achieved and in the assessment task to verify that the outcome has in fact been achieved" (2007, p. 52). The focus is thus squarely on what students must learn, what they are to do with it, and how they are to learn it, not on what topics teachers are to teach. Seen in this framework, the translation task carried out in conditions mirroring those of professional work is au authentic means of achieving learning outcomes, as has been ably demonstrated by Kiraly (2000) and others in their applications of constructive alignment to translation teaching.

For Wiggins, "achievement targets" are the intended learning outcomes of a course or program. The challenge is to formulate targets in measurable terms so that, as assessors, teachers know what evidence to look for in determining how well students have performed. For example, Kelly proposes the following overall outcome for a translation program:

> On completion of the course, students will have acquired the necessary competences (knowledge, skills and attitudes) to be able to join the translation profession in any of its specialized areas in this country or abroad at a junior level. (2005, p. 36)

As a broad program outcome, the statement works. However, outcomes for individual courses designed to develop one or more specific competencies should be expressed in more precise terms if they are to serve as helpful guidelines for teacher and student. Thus a statement of an intended learning outcome or achievement target statement such as "At the end of the course, students will be able to translate texts well" is vague and unlikely to yield useful measurements. What kinds of texts will they be translating? What does "translate well" entail? The outcome statement must be supported by precise information to specify levels of achievement of the outcome.

3.2 Intended learning outcomes, translation competencies and types of knowledge

In a professional discipline such as translation, and given the rich, authoritative literature on translation competencies, achievement targets (learning outcomes) may be usefully expressed in terms of those competencies. Many have focused on differentiating translation and translation studies from foreign language learning and bilingualism by isolating a translation-specific competency or skill that would be required of students in addition to linguistic knowledge, general knowledge and psychological competencies such as motivation (Roberts, 1984; Neubert, 2000; Colina, 2003; Pym, 2003; Kelly, 2005; Angelelli, 2009; etc.). For the purposes of this article, we will refer to six competencies: translational, linguistic (knowledge of SL and TL), textual (knowledge of textual and rhetorical features), cultural/encyclopaedic, reasoning, and strategic (research, monitoring, and decision-making skills).

Acquiring competencies and achieving outcomes means acquiring and demonstrating knowledge. Educational theorists from Bloom on have categorized learning achievements and outcomes in terms of types of knowledge. In their revision of Bloom's taxonomy of education objectives, Anderson and Krathwohl (2001) propose four types of knowledge (factual, conceptual, procedural and metacognitive). Biggs and Tang distinguish between declarative knowledge (knowing-what) and functioning knowledge (knowing-how) (2007, 72). The latter is based on the idea of successful performances of various kinds and requires a solid foundation of the former (factual and conceptual knowledge). Leinhardt et al. (1995) and Bromme and Tillema (1995) make a similar distinction between academic knowledge acquired at university (declarative, abstract and conceptual) and professional knowledge acquired in practice (procedural, "how-to", specific and pragmatic). Translation training illustrates the need for all these types of knowledge. Students must develop their knowledge of discrete concepts, terms, typographical and grammar rules, etc. (declarative and conceptual knowledge), and they need to acquire methods for conducting research and producing translations (procedural or functioning knowledge). They also need to develop metacognitive knowledge to hone their strategic competency and, in particular, develop monitoring skills to revise and improve their own work.

Clear statements of intended learning outcomes for a course, the types of knowledge and competencies being targeted, and the degrees of knowledge or competency to be achieved will be integral to the student assessment model, providing valuable information for a variety of stakeholders, including the students themselves, colleagues, and even prospective employers. The student, the teacher and the assessor have to know "what aspects of an individual's translation ability should be assessed" (Angelelli, 2009, 15), and any test, examination or assignment must be built in such a way as to yield adequate information on students' degree of success in demonstrating acquisition of the target competencies and knowledge.

3.3 Standards model of assessment

Once we set an achievement target or intended learning outcome and combine it with a statement of types and degrees of target competencies and knowledge, we are at the same time setting a standard, and it is here that the value of the linkage between "learning" and "doing" in the constructive alignment approach becomes very clear: "Say what you want students to be able to do, teach them to do it and then see if they can, in fact, do it. […] The alignment is perfect" (Biggs and Tang, p. 177). In a translation course, the teacher sets the standard, or set of standards, and then sees whether students can achieve the standard by measuring translation performance against specific assessment criteria. The standards model of assessment is thus criterion-referenced.

3.4 Authentic translation quality assessment and standards

Wiggins stresses the importance of moving away from evaluating content knowledge alone and, instead, of assessing "students' ability to perform on complex tasks" (1998, p. 15). The assessment task must be designed to elicit from the student a performance reflecting "the kind of understanding that requires an *active demonstration* of the knowledge in question, as opposed to talking or writing about it. This is referred to as 'authentic assessment'" (Biggs and Tang, p. 181).

In a specialized translation course, a translation project aligned with the content of learning activities and exemplifying the work done by professional translators in the area or areas of specialization concerned should serve adequately as an authentic translation and assessment task. It remains for us to establish the targets, standards, criteria and indicators needed to ensure a valid and reliable assessment. Given the professional thrust of university translation programs and the focus that the educational theorists cited in this article have placed on the "active demonstration" of knowledge in practice, such requirements should logically be based on the standards of the translation profession.

Therefore, in developing our model, we should ask questions directly related to the competencies and performances reflected in professional standards. For example, how accurate and clear do students' translations have to be? How correct, idiomatic and readable to do they have to be? After all, those standards are the targets students will have to meet eventually in order to be successful employees or entrepreneurs. The Translation Studies and grey literature abounds in professional standards of various stripes against which the work of professional translators can be measured. There is also no shortage of examination rating schemes against which candidates for national or regional certification are evaluated.

We will refer to two such schemes: those of the Canadian Translators, Terminologists and Interpreters Council (CTTIC) and the American Translators Association (ATA).

The CTTIC marker's and candidate's guides highlight the following statement:

> A candidate is judged *competent* if the translation provided is *faithful* and *idiomatic* and *requires little or no revision*. (CTTIC, March 2009)

In fact, as a standard, the statement "the translation is faithful and idiomatic and requires little or no revision" conflates at least two translation competencies identified by theorists — translational and linguistic — and implies all four knowledge dimensions (declarative, conceptual knowledge, procedural, and metacognitive). In practice, a candidate demonstrates overall competence by translating two texts and scoring an average of 70% or higher. For each translation, each error results in a deduction of 1 (e.g. for a typographical error) to 10 points (e.g. for a major error of transfer), depending the type and seriousness of the error.

The purpose of the ATA certification examination is defined as follows:

> The ATA certification examination tests for professional translation skills. It is designed to determine whether a candidate is able to produce a translation that is professionally usable within the framework provided by the Translation Instructions. The skills in question are defined by the positive answers to four broad questions:
>
> *Does the translation demonstrate compliance with the specifications of the Translation Instructions?*
>
> *Does the translation demonstrate understanding of the overall content, purpose, and argument of the examination passage?*
>
> *Does the translation demonstrate competent familiarity with translation strategies of various kinds?*
>
> *Does the translation demonstrate good writing in the target language?*

Thus the ATA examination is designed to assess knowledge and competencies at a finer-grained level than that of CTTIC: ability to comply with specifications, ability to understand overall content, purpose and argument of source text, strategic competency, and linguistic competency. In addition, ATA provides graders with a "Rubric for Grading." While evaluating a translation by means of a points deduction scheme, the grader is also asked to evaluate the candidate's translation performance in terms of four dimensions — *usefulness/transfer*, *terminology/style*, *idiomatic writing*, and *target mechanics* — and assign to it one of five grades for each dimension: *standard, strong, acceptable, deficient* and *minimal*. There is a description of each grade across all four dimensions. In addition, the grader is asked to measure the translation "against

the ideal performance," defined for each dimension in a "standard" grade. The descriptors serve to report on the quality of the translation itself. Here are the descriptors for the "Standard" and "Acceptable" grades:

Usefulness/transfer

Standard: The translated text is fully usable for the purpose specified in the Translation Instructions. The meaning and sense of the source text have been fully and appropriately transferred to the translated text. .

Acceptable: Translated text transfers meaning in a manner sufficiently consistent with the Translation Instructions. Translation contains occasional and/or minor transfer errors that slightly obscure or change meaning.

Terminology/style

Standard: Terminology is appropriate in context. Style and register are appropriate for the topic in the target language and for the specified audience.

Acceptable: The Translated text contains occasional and/or minor inappropriate term or style/register choices. Such errors may slightly obscure meaning.

Idiomatic writing

Standard: Translated text reads smoothly. Wording is idiomatic and appropriate for the topic in the target language and for the specified audience.

Acceptable: Translated text contains occasional unidiomatic or inappropriate wording. Such errors may slightly obscure meaning.
.

Target mechanics

Standard: Translated text fully follows the rules and conventions of target language mechanics (spelling, grammar, punctuation, etc.).

Acceptable: Translated text contains occasional errors in target language mechanics. (ATA, 2013)

The overall *usability* of the translation is established in a broader definition of each grade and is related to a specific level of performance:

Strong: The target text would require little if any editing in order to be used for the purpose specified in the Translation Instructions.

Acceptable: A client requesting this translation could use the text for the purpose given in the Translation Instructions after some work by a bilingual editor and/or a target language copyeditor. (ATA, 2013)

In short, the ATA provides extensive documentation to describe and define a required level of competency and, at the same time, a standard or set of standards in terms of the quality of the text itself. In addition, the perspective is clearly functionalist, with frequent references to translation use and purpose and target audience. Like CTTIC, the ATA too combines its descriptors with a quantitative grid. Candidates translate

two texts of 150–200 words, and markers deduct points for errors made to generate a percentage mark. Serious errors do not necessarily result in a "fail." In the end, an "Acceptable" grade, representing a "level of obvious competence with some room for growth," meets the requirements for certification, just as the CTTIC pass mark of 70% allows for a number of major errors.

The CTTIC scheme lacks any definition of specific competencies or criteria, while the ATA breakdown of dimensions is arguable — particularly its division of target language competency into three dimensions. Indeed, Angelelli takes the ATA to task for excluding macrotextual elements, text function and the needs of the target readership from its model (Angelelli, pp. 29-30). That being said, they are in essence national professional translation standards, contain statements of criteria against which professional translators and translations are to be judged, and therefore provide an authoritative starting point for developing a similar criterion-referenced framework for student assessment. To be sure, in developing a framework for student assessment, we will need to factor in a slower production rate and the need to set tasks of lower difficulty and complexity than those alluded to the ATA performance level definition. That being said, translation students are being trained to become, within a relatively short time, useful, cost-effective resources in a highly competitive environment. It follows that the achievement targets set for them should not be too far removed from professional standards and that the highest performance targets should mirror professional excellence.

3.5 Criteria

Assessment involves three steps: (1) setting the criteria; (2) selecting relevant evidence (indicators) for making a valid judgment against the criteria; (3) making a judgment about the degree to which the criteria have been met and assigning a corresponding performance level or grade. We have referred above to the "criteria" in the ATA grading documents, but we need to define what the term means exactly. An assessment process should tell us whether and how well a student can achieve the intended learning outcomes and demonstrate the target competencies, and measurement of degree of achievement requires a set of criteria. Wiggins defines a criterion in terms of performance:

> A criterion is a way of defining success at meeting a target achievement or educational outcome….When criteria are met, then we may conclude not only that the specific performance has been successful…but that a larger educational goals has been addressed. Criteria are always necessary and sufficient: the larger success has not occurred unless the criteria are met, and the criteria apply in all attempted performances of the same type. (1998, pp. 128-29).

Thus the proposed framework must enable the teacher/assessor to judge student performance against criteria. The assessor no longer merely asks how many marks he

or she should give a particular component of the performance (under a purely quantitative measurement model); the assessor also asks how well the performance meets the criteria for a given level of quality or competency.

3.6 Indicators

However, in order for the judgment to be valid and not "impressionistic," the framework needs to include explicit evidence, called indicators, showing that the criteria have been met and serving to justify the eventual grade or rating. According to Wiggins:

> An indicator is a behavior or trait that is typical of a particular performance being assessed. It is a concrete sign or symptom of a criterion being met and thus makes assessment and self-assessment easier" (1998, p. 129).

By extension, it also provides feedback on whether, and to what degree, the student has met the standards set. In a translation context, such signs are the presence of strong points and weak points in a student's performance and may be expressed in terms of the types and quantity of defects detected and effective stylistic, structural and transfer devices used. Thus while Wiggins and Biggs and Tang argue for a holistic instead of a measurement-based, or quantitative, approach to assessment, this does not mean that the assessor must eschew the analysis of details. This is especially the case in translation, where microtextual defects affect macrotextual quality, and vice-versa.

3.7 The rubric

Having fleshed out the various components of the educational assessment model, we must now "put it all together" and develop a scoring rubric for a translation course or program. The rubric "tells potential performers and judges what elements of performance matter most and how the work is to be distinguished in terms of relative quality" (Wiggins, 1998, p. 153). In other words, the goal is to develop criteria and indicators along a spectrum of performance levels and thus provide a set of descriptors or guidelines for each level (or grade) of performance. In principle, this roll-up of assessment components should provide sufficiently detailed information to ensure a measure of consistency and greater inter-rater reliability in determining the value of university grades.

A complicating factor in the construction of rubrics is that the educative assessment model is criterion-referenced: the judge is assessing performance against a number of discrete criteria, such as textual adequacy and quality of content, and is therefore working initially with a componential or "trait-analytic" rubric (Wiggins, 1998, p. 153). However, Wiggins does propose a procedure for moving from a componential to a holistic rubric, which we will exploit here.

3.8. Grade nomenclature

It is helpful to both assessor and assessed if an evaluative descriptor or label is associated with each letter or numerical grade. The "excellent–very good–good–fair–poor" generic breakdown is commonplace and does serve to situate the quality of performance within a range of grades or standards. Colina's CFA "Assessment Summary and Recommendations" sheet (2009, Appendix 1) is particularly useful because it contextualizes the grades and give them more precise and informative labels indicating the relative usability of a translation for publication purposes (see 5.6 below). An argumentation-centred assessment rubric also presents a gradation based on usability for specific purposes: publication, information, minimum and substandard (Williams, 2004, pp. 145-47). The differences between the four grades reside in absence or presence of critical and major defects in conveying the ST author's reasoning and arguments (and thus in conveying the purpose of the translation) and in meeting other requirements relating to end use.

An educational assessment rubric should be equally informative, telling translation students how much progress they have made toward their ultimate goal of professional-level translation competency. In the case of student performance, Wiggins stresses the importance of authentic tasks mirroring real professional work, and the goal of authentic performance in an academic environment may very well be to produce work of publishable quality or information quality with due regard for source text difficulty and real-world time constraints. What is of equally informative value is the student's progress toward translation competency, which can be defined, mutatis mutandis, in terms of the four CFA components and can be exploited to yield a rubric for a specialized translation course.

4. The functional-componential approach

We can now start to draw parallels between educational assessment principles, professional standards, and the componential-functional approach (CFA) proposed by Colina for the assessment of professional translations.

4.1 A criterion-referenced approach

Theoretical models, Colina contends, focus on only some aspects of translation, and some of them "overlook the fact that quality in translation is a multifaceted reality, and that a general comprehensive approach to evaluation may need to address multiple components of quality simultaneously" (2009, p. 239).

Accordingly, she proposes what she considers to be a TQA approach based on theory (functionalist and textual models of translation) but also applicable in professional and

educational settings. In the CFA model, the criteria for specialized translation assessment relate to four "components": functional adequacy of the translation, quality of content (transfer), quality of target language, and quality of specialized content. The criteria to be met for a translation to be assigned the highest grade ("Publish and/or use as is") are as follows:

> Functional adequacy (and impact of performance): *The translation achieves its intended purpose and would achieve the intended effect on the reader.*
> Quality of content: *The translation faithfully renders the arguments, reasoning and details of the source text (coherence and attention to detail).*
> Textual adequacy: *The translation is idiomatic, readable and typographically and idiomatically correct.*
> Quality of specialized content: *The translator uses terminology from authoritative, relevant sources and specialized concepts are faithfully conveyed.* (Colina, 2009, pp. 259–260)

The range of criteria would seem to be comprehensive, similar to those of the professional associations, and therefore adaptable to a student assessment framework designed to relate students' performances to that of professionals.

In the componential-functionalist approach, multiple raters receive a translation brief, in which the translation requestor specifies, among other things, the purpose of the text, the target readership, and his or her priorities among the four TQA components. The approach is thus based, at least in part, on a "user-defined notion of quality" (2009, p. 240) in that the relative weights given to the four components necessarily reflect the requestor's order of priority.

No points are deducted for specific shortcomings in a translation or for a given number of errors. Having analyzed the translation, the rater selects one of four quality levels for each component. Each quality level is associated with a description listing the features of a text corresponding to the level concerned. Colina's approach is quantitative only to the extent that numerical values (percentage weights) are attached to each quality level, again reflecting the requestor's order of priority among the components. Below are the "Rating Instructions" for the *target language* component (2009, p. 259).

Target Language

Category Number	Description	Check one box
1.a	The translation reveals serious language issues. Ungrammatical use of the target language, spelling mistakes. The translation is written in some kind of "third language" (neither the source nor the target language). The structure of source language dominates to the extent that it cannot be considered a sample of target language text. The amount of transfer from the source cannot be	

		justified by the purpose of the translation. The text is extremely difficult to read, bordering on being incomprehensible.	
1.b		The text contains some unnecessary transfer of elements/structure from the source text. The structure of the source language shows up in the translation and affects its readability. The text is hard to comprehend.	
1.c		Although the target text is generally readable, there are problems and awkward expressions resulting, in most cases, from unnecessary transfer from the source text.	
1.d		The translated text reads similarly to texts originally written in the target language that respond to the same purpose, audience and text type as those specified for the translation in the brief. Problems/awkward expression are minimal is existent at all.	

In the Instructions, the rater is asked to check the description that best fits the text for each category. The description is, in fact, a set of indicators of the quality level, much like the ATA Rubric for Grading. The rater can add comments and examples to justify his or her rating. The rater then fills out a scoring worksheet for each component, assigning the translation a score based reflecting the priority assigned to that component by the requestor and an "assessment summary and recommendation" on what should be done with the translation. The options range from "Publish and/or use as is" to "Redo translation."

4.2 Strengths and weaknesses of the componential-functionalist approach

In my view, the strengths of CFA are as follows:

1) It is criterion-referenced, as evidenced by the descriptors, or sets of indicators, of the four quality levels for each TQA component.
2) It is flexible and adaptable in that the criteria can be adjusted and the four components can be given different relative weights in light of user priorities, the purpose of the translation, the prospective readers, and end use.
3) While not based on errors counts, it nonetheless includes a quantitative dimension (weightings and percentages), without which it would be impossible to arrive at a decision on the usability of the translation.
4) Because CFA covers four core features of translations and because it weights those features, it can provide for componential and overall assessments.
5) The four TQA components are broad enough to cover all quality features, so CFA meets the requirements for validity. It is appropriately designed to assess what it is supposed to assess: translation quality.
6) CFA is based on a proven theoretical framework, functionalism.
7) While the primary purpose of CFA as a TQA tool is to measure usability, its comprehensive coverage of quality features and the recommendation scale suggest that it could be the starting point for a student translation assessment model.

Now, what about the possible shortcomings of CFA is applied in an education context?

- The use of multiple raters to ensure reliability is costly and time-consuming. In an educational context, the teacher/professor tends to be the sole rater.
- The reliance on the user or customer to determine which TQA components are most important and should therefore be more heavily weighted could make for unstable TQA results. In a translator training context, the teacher/professor as a translation expert would determine the relative values of components in light of the objectives of the course or program concerned.
- Although the tally sheet assigns an overall percentage to the translation, there is no quantitative (numerical) or descriptive definition of each overall rating. Instead, that rating takes the form of a recommendation to publish or use the translation as is, to make minor changes, etc.
- There is no reference to the concepts of critical, major and minor error (or defect), which are core characteristics of quality control and assurance processes in industry.

The educational TQA model proposed below is designed to resolve these issues.

5. A holistic-componential model for translation student assessment

In light of the preceding discussion, the proposed model for the assessment of student translations in a specialized translation course is designed to
- be based on an authentic assessment construct, or set of target professional competencies;
- provide for assessment against criteria mirroring progress toward professional standards;
- be functionalist, providing for assessment, to the extent possible, in light of the purpose of translation and target readership needs;
- be criterion-referenced and therefore componential, providing for assessment against specific professional competencies/types of knowledge;
- be holistic, providing for an overall assessment of ability and performance that goes beyond (but does not exclude) error counting;
- reflect the principles and requirements of education theory and include a statement of intended learning outcomes as well as a rubric comprising a set of target competencies/knowledge types, a range of grades representing different levels of acquisition/demonstration of those competencies, and criteria and indicators defining the competency levels;
- provide useful information to teachers, students, and other stakeholders;
- ensure validity by generating appropriate, meaningful, and useful inferences about student competencies from the assessment;

- facilitate grading and help justify the grades given by assigning numerical values to student performance on each component (competency) and generating an overall (holistic) assessment from those values.

5.1 Statement of intended learning outcomes

Biggs and Tang recommend that the number of outcomes be limited. The key is that they reflect the target competencies and types of knowledge.

1. Translate documents in human resources management (statistics, etc.) to the standard of accuracy required of an intermediate-level trainee by ensuring that the function and informative intent, and the reasoning and argumentation of documents is fully and effectively communicated. Target competencies: Translational, linguistic, textual, cultural/encyclopaedic, reasoning, strategic. Target knowledge dimensions: Declarative, conceptual, procedural, metacognitive.
2. Edit and revise their own work in specific fields to produce readable, idiomatic and typographically grammatically correct documents and that target language textual features are used appropriately. Target competencies: Linguistic, textual, and strategic. Target knowledge dimensions: Declarative, conceptual, procedural, and metacognitive.
3. Apply specific methods and techniques effectively and efficiently to the translation of specialized texts. Target competencies: Reasoning, strategic. Target knowledge dimensions: Procedural, metacognitive.
4. Understand specialized concepts, retrieve correct terminology for those concepts by accessing authoritative resources, and explain and justify decisions and choices adequately. Target competencies: Linguistic, cultural/encyclopaedic, reasoning, strategic. Target knowledge dimensions: Declarative, conceptual, procedural, metacognitive.

As Angelelli points out, due to testing formats and constraints and technology limitations, it may not be possible to assess explicitly the full range of outcomes and, therefore the full range of competencies, in every assessment task (2009, p. 37). For example, achievement of ILO 3 (which Angelelli and others would consider a strategic competency) "is only evident in its effect: strategic competence is truly demonstrated in the absence of problematic translations in the final product" (2009, p. 37).

Having established the ILOs for the course, we can now construct a rubric with specific criteria and indicators to describe varying degrees of competency. Because we want to align the rubric with the functionalist approach and its focus on usability and relate student performance squarely to professional performance, we draw extensively on the components and terminology of CFA and the ATA descriptors. In the interest of (relative) simplicity, we have assigned the same criteria and indicators to all grades within the same alpha range (D combined with D+, etc.). The grades could be

differentiated if need be. In addition, we restrict the number of components to three (we have excluded the "Non-specialized content-meaning" component in CFA) in order to minimize, to the extent possible, repeated assessment of the same items in two or more components.

5.2 Assessment rubric

Component 1 (ILOs 1, 3 & 4): Transfer/functional and textual adequacy
Competencies: Translational, linguistic, textual, cultural/encyclopaedic, reasoning, strategic
Weighted value: 40

Grade	Description Criteria: accuracy, coherence, quality of reasoning/argumentation, consideration of text function (informative, explanatory, directive, persuasive) and target readership expectations and requirements	Value range	Score
A-, A & A+ **Very high level of competency** **Very high degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: Translation accurately reflects meaning of ST, without unwarranted alterations, omissions or additions. Nuances or shades of meaning have been accurately rendered. Translation performs intended function and meets target readership expectation/requirements. Demonstration of very high degree of declarative and conceptual knowledge of SL, TL, of procedural knowledge of translation methods, and of monitoring for error.	**32-40**	
B & B+ **High level of competency High degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: Minor alterations in meaning, additions or omissions. Translation generally performs intended function and meets target readership expectation/requirements. Demonstration of high degree of knowledge of SL, TL, translation methods, and monitoring.	**28-31**	
C & C+ **Emerging**	Indicators: Some unjustified changes in meaning,	**24-27**	

Grade	Description	Value range	Score
competency **Moderate degree of declarative, conceptual, procedural, and metacognitive knowledge**	omissions and/or additions, some of them significant. Translation shows some consideration of intended function and target readership expectation/requirements, but, for example, elements of argumentation and TL textual features are not rendered. Demonstration of moderate knowledge of SL, TL, translation methods, and monitoring.		
D & D+ **Marginal competency** **Limited degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: Several deviations from ST, some of them significant. Translation shows limited consideration of intended function and target readership expectation/requirements, and several elements of argumentation and TL textual features are not rendered. Demonstration of limited knowledge of SL, TL, translation methods, and monitoring.	**20-23**	
F **Competency not demonstrated** **Very low degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: Many unwarranted and significant deviations from ST. Inaccurate renderings and/or important omission and additions. Translation shows very limited consideration of intended function and target readership expectation/requirements. Demonstration of very defective knowledge of SL, TL, translation methods, and monitoring.	**0-19**	

Component 2 (ILOs 2 and 3): Target language **Competencies: Linguistic, textual, reasoning, strategic** **Weighted value: 40**			
Grade	**Description** Criteria: quality of typography, grammar and usage, readability, idiomaticity, cohesion, consideration of text function (informative, explanatory, directive, persuasive) and target readership expectations and requirements	**Value range**	**Score**

A-, A & A+ **Very high level of competency** **Very high degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: Virtually no TL errors. The text reads as if it was originated in TL and meets target readership's expectations regarding text function and genre. \very few, if any, typographical, grammatical and usage errors. Demonstration of very high degree of knowledge of TL and TL textual/rhetorical features and conventions and of monitoring for error.	**32-40**	
B & B+ **High level of competency** **High degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: A few minor TL errors. The text generally reads as if it was originated in TL and generally meets target readership's expectations regarding text function and genre, but there some awkward expressions and calques and occasional typographical, grammatical and usage errors. Demonstration of high degree of knowledge of TL and TL textual/rhetorical features and conventions and monitoring for error.	**28-31**	
C & C+ **Emerging competency** **Moderate degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: A number of typographical, grammatical and/or usage errors, some of them significant. SL typography, grammar, lexicon and usage shows up in the translation and adversely affects readability. Cohesion between propositions is sometimes defective. Several typographical, grammatical and usage errors. Demonstration of moderate degree of knowledge of TL and TL textual/rhetorical features and conventions and monitoring for error.	**24-27**	
D & D+ **Marginal competency** **Limited degree of declarative, conceptual,**	Indicators: Many typographical, grammatical and/or usage errors, some of them major. SL and rudimentary typography, grammar, lexicon and usage show up in the translation in several instances and	**20-23**	

procedural, and metacognitive knowledge	affects readability significantly. Cohesion between propositions is often defective. Demonstration of limited knowledge of TL and TL textual/rhetorical features and conventions and of monitoring for error.		
F **Competency not demonstrated** **Very low degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: Too many typographical, grammatical and/or usage errors, a number of them major. SL typography, grammar, lexicon and usage dominate and adversely affect readability through the translation. In no way is it an example of an acceptable TL text Cohesion between propositions is highly defective. Demonstration of very low degree of knowledge of TL and TL textual/rhetorical features and conventions and monitoring for error.	**0-19**	

Component 3 (ILOs 3 and 4): Terminology/research Competencies: Reasoning, strategic Weighted value: 20			
Grade	**Description** Criteria: quality of terminological, authoritativeness of sources, adequacy of explanations and justifications for choices.	**Value range**	**Score**
A-, A & A+ **Very high level of competency** **Very high degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: Terms are accurate and appropriate to field. Explanations are complete and valid. All sources are appropriate. Demonstration of very high degree of (declarative) knowledge of specialized terminology and (metacognitive) knowledge of means of evaluating of sources.	**16-20**	
B & B+ **High level of competency** **High degree of declarative, conceptual, procedural, and**	Indicators: Terms are generally accurate and appropriate. Explanations and sources are generally of good quality. Demonstration of high degree of (declarative) knowledge of specialized terminology and (metacognitive)	**14-15**	

metacognitive knowledge	knowledge of means of evaluating of sources.		
C & C+ **Emerging competency** **Moderate degree of declarative, conceptual, procedural, and metacognitive knowledge**	Indicators: Several terminological errors affecting terminological content. Some important explanations of translation decisions are missing and/or some sources are of poor quality or are not given. Demonstration of moderate degree of (declarative) knowledge of specialized terminology and (metacognitive) knowledge of means of evaluating of sources.	**12-13**	
D & D+ **Marginal competency** **Limited degree of declarative, conceptual, procedural, and metacognitive knowledge**	Serious and frequent errors in terminology and/or specialized content. Limited explanations of translation decisions and/or use or several unreliable sources. Demonstration of limited degree of (declarative) knowledge of specialized terminology and (metacognitive) knowledge of means of evaluating of sources.	**10-11**	
F **Competency not demonstrated** **Very low degree of declarative, procedural and metacognitive knowledge**	Accurate, appropriate terms not used. Insufficient explanations of translation decisions and/or poor choice of sources. Demonstration of very low degree of (declarative) knowledge of specialized terminology and (metacognitive) knowledge of means of evaluating of sources.	**0-9**	

5.3 From componential to holistic assessment

The next step is to derive an overall score (percentage) by adding up the scores assigned to the four weighted components and presenting the result in the "Tally Sheet" below, modeled closely on that of Colina. Once the percentage is known, the assessor checks the appropriate box in the "Holistic Assessment and Grading Summary" to indicate the grade assigned. The Summary includes an interpretation of the grade stating how far the student has progressed toward the target of professional-quality translation, based on the assessment.

Tally Sheet		
Component	**Category Rating**	**Score Value**

Transfer/functional and textual adequacy	40	
Target language	40	
Terminology/research	20	
Total Score	100	%

Holistic assessment and grading summary

Percentage	Check one	Grade	Description
80-100		A-, A, A+	Very high level of competency/significant progress toward professional knowledge in all dimensions. Demonstrated ability to produce work of professional quality requiring only minor revisions.
70-79		B, B+	Demonstrated competency/clear progress toward professional knowledge in all dimensions and ability to produce work adequate for information purposes or of publishable quality after some revision.
60-69		C, C+	Emerging competency. Demonstrated potential for producing translations that are usable after extensive revision, but significant progress required in one or more competencies/knowledge dimensions.
50-59		D, D+	Marginal competency. Given the number and seriousness of errors, student shows limited potential for producing usable translations. Significant progress required in all competencies/knowledge dimensions.
0-49		F	Competency/knowledge not demonstrated. Given the number and seriousness of errors, the student does not show potential for producing professional-quality work over the medium or long term.

6. CONCLUSION

According to Biggs and Tang, a valid assessment must be of the student's total performance, but at the same time the conceptual framework underlying assessment must relate the whole to its parts (2007, pp. 184-85). By establishing a comprehensive set of quality components and criteria and associating with them specific competencies, types of knowledge and indicators, the proposed model can generate an assessment of overall performance and competency (holistic assessment) from an

assessment of performance against specific criteria (componential assessment).

Furthermore, the model does not abandon the quantitative dimension of assessment. It combines it with the qualitative dimension by providing for a qualitative assessment of each of the three selected components and, at the same time, including consideration of the number and seriousness of defects in the calculation of component scores and final percentage. The quantitative dimension always facilitates reporting and justification of grades.

Any assessment must prove its validity and reliability. So is this particular model valid in that it generates appropriate, meaningful, and useful inferences about student competencies? Its validity is based on two factors: (1) the alignment of the criteria, indicators and grade definitions with the intended learning outcomes, which helps to ensure that the model assesses what it is designed to assess; (2) the level of detail in the indicator and grade descriptors, which provides the various actors with useful information about the results of the assessment.

Is it reliable? Yes, it is, for the same reasons. The alignment of assessment factors with achievement targets and the level of detail and quantitative elements in the descriptors and grade scores should limit the risk of inconsistency in grading.

Wiggins recommends that teacher/assessors use grades that "stand for something clear, stable and valid," that are "linked to credible and important…standards for assessing performance on authentic tasks," and that can serve to "measurably improve student performance over time so that standards once thought very high and reachable by only a few become reasonable expectations for many students" (Wiggins, 1998: 12). The proposed holistic-componential model aligns learning and assessment tasks with authentic professional quality standards and, at the same time, incorporates a range of grades offering "potential for growth" and, assuming students act on the information in the rubric and summary, a pathway toward professional competency and certification.

Bibliography:
American Translators Association. Website accessed February 9, 2013: http://www.atanet.org/certification/aboutexams_overview.php)

Anderson, L., and D. R. Krathwohl (Eds.). (2001). A *Taxonomy for Learning, Teaching, and Assessing*. New York: Longman.

Angelelli, C. (2009). Using a Rubric to Assess Translation Ability. In C. Angelelli and H. Jacobson (Eds.). *Testing and Assessment in Translation and Interpreting Studies*. John Benjamins: Philadelphia, 13–47.

Anckaert, P., J. Eyckmans, and W. Segers. Pour une évaluation normative de la compétence en traduction. (2008). *ITL International Journal of Applied Linguistics*, 155: 53–76.

Biggs, J., and C. Tang. (2007). *Teaching for Quality Learning at University*. 3rd ed. Maidenhead: Open University Press.

Bloom, B. (1956). *Taxonomy of Educational Objectives.* Boston: Allyn and Bacon.

Bromme, R, and H. Tillema. (1995). Fusing Experience and Theory: The Structure of Professional Knowledge. *Learning and Instruction*, 5: 261–267.

Colina, S. (2003). *Teaching Translation: From Research to the Classroom.* Boston: McGraw Hill.

———. (2008). Translation Quality Evaluation: Empirical Evidence for a Functionalist Approach. *The Translator*, 14, 1: 97–134.

———. Further Evidence for a Functionalist Approach to Translation Quality Evaluation. (2009). *Target*, 21, 2: 235–264.

Delisle, J. (1980). *L'analyse du discours comme méthode de traduction*. Ottawa: University of Ottawa Press.

Eyckmans, J., P. Anckaert, and W. Segers. The Perks of Norm-referenced Translation Evaluation. (2009). In C. Angelelli and H. Jacobson (Eds.). *Testing and Assessment in Translation and Interpreting Studies*. John Benjamins: Philadelphia, 73–92.

Kelly, D. (2005). *A Handbook for Translators: A Guide to Reflective Practice*. Manchester, UK: St. Jerome Publishing.

Kim, R. (2006). Use of Extralinguistic Knowledge in Translation. *Meta*, 51(2): 284-303.

Kiraly, D. (2005). *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*. Manchester: St. Jerome Publications, 2000.

Larose, R. (1998). Méthodologie de l'évaluation des traductions. *Meta*, 43, 2: 163-186.

Leinhardt, G., K. McCarthy Young, and J. Merriman. (1995). Integrating Professional Knowledge: the Theory of Practice and the Practice of Theory. *Learning and Instruction,* 5: 401–408.

Lentz, L., and J. Hulst. (2000). Babel in Document Design: The Evaluation of Multilingual Texts. *IEEE Transactions in Professional Communication*, 43, 3: 313–322. Martinez Melis, N., & Hurtado, A. (2001). Assessment in Translation Studies: Research Needs. *Meta*, 46 (2): 272-87.

Neubert, A. (2000). Competence in Language, in Languages, and in Translation. In C. Schäffner and B. Adab (Eds.), *Developing Translation Competence*. Philadelphia: John Benjamins, 3–18.

Pym, A. (2003). Redefining Translation Competence in an Electronic Age. In Defence of a Minimalist Approach. *Meta*, 48(4): 481-497.

Roberts, R. (1984). Compétence du nouveau diplômé en traduction. In Société des traducteurs du Québec / Conseil de la langue française *Traduction et qualité de langue.* Québec: Conseil de la langue française), 172-77.

Translation Automation User Society (TAUS). (2011). Enabling Better Translation. http://www.translationautomation.com/

Vienne, J. (1998). Teaching What They Didn't Learn as Language Students. In K. Malmkjaer (Ed.). *Translation and Language Teaching. Language Teaching and Translation*. Manchester: St. Jerome Publishing: 111–116.

Wiggins, G. (1998). *Educative Assessment. Designing Assessments to Inform and Improve Student Performance*. San Francisco: John Wiley and Sons.

———, and J. McTighe. (2005). *Understanding by Design*. 2nd ed. Alexandra, VA: ASCD.

Williams, M. (2001). Application of Argumentation Theory to Translation Quality Assessment. *Meta*, 46(2): 326-344.

————. (2004). *Translation Quality Assessment: An Argumentation-centred Approach*. Ottawa: University of Ottawa Press.