Nordic Journal of Language Teaching and Learning

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

# Intended or unintended consequences? Investigating the ways in which the National Tests of English are implemented and acted on in Norwegian schools

***Craig Grocott***
*University of Bergen*

## Abstract

*This article reports on an investigation into the role the National Tests of English (NTE) and their results play in Norwegian eighth-grade classrooms. Previous and current opposition to the tests from some teachers and pupils gave rise to the question of whether the tests are being used as recommended, based on documents produced by the Directorate of Education and Training. The study proceeded from the premise that consequential validity (Messick, 1994) could be under threat in cases of clear discrepancies between intended and actual consequences and uses of the tests and their results. A mixed methods study was conducted among eighth-grade teachers of English, consisting of a quantitative digital survey and qualitative, semi-structured interviews. In total, 43 English teachers participated in the study.*

*Results indicated a lack of uniformity in the uses of the NTE, with both the nature and levels of engagement in individual schools being determined by factors such as principals' concerns, time constraints, parents' interest levels and teachers' own views on the usefulness of the tests and the results. Validity is threatened when unintended consequences take the place of intended consequences (Chalhoub-Deville, 2015). The study reveals that, in around half of the schools involved, this seems to be happening, partly ascribable to a lack of time and/or interest in the tests. As respondents reported that schools' allocations of time and resources to the tests are largely determined by school principals, a follow-up study with principals is recommended.*

**Keywords:** National tests, validity, consequential validity, English language testing

## Introduction

All large-scale language assessments, whether high or low stakes, have a purpose and, inevitably, a set of consequences. It is important to ascertain whether this set of consequences corresponds to the intended uses of the assessment in question (Im, Shin & Cheng, 2019). The National Tests of English (NTE) are such tests with a clearly stated purpose, and there is therefore a clear need to assess the manifestation of their consequences. The NTE are digital English reading tests that have been taken by virtually all fifth and eighth-grade pupils in Norway since 2004, and their overall purpose is to provide teachers, school leaders, parents, and other local and national stakeholders with information about pupils' English reading ability (Utdanningsdirektoratet, 2019). This information is intended to be a starting point for formative assessment and quality development, which underlines the importance of the

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

consequences of the information. The present study assesses the uses and consequences of the NTE within a validity framework, with the idea that validity can be threatened if there is a clear discrepancy between the intended uses of a test and what happens in schools. The framework is partly based on Messick's (1996) presentation of consequential validity. Messick discusses the relationship between the stated goals of tests and the consequences that arrive from them. This is expanded upon in the theoretical background section.

The present study is part of a wider project, focusing on various validity aspects of the eighth-grade NTE. The wider project addresses validity aspects such as content, developed *a priori* by test makers. The present study however concerns itself with the *a posteriori* relationship between stakeholders and the NTE, especially in terms of uses and interpretations. The eighth-grade tests are the focus of the wider project as they test a wider range of skills than the fifth-grade tests and are graded across five mastery levels, as opposed to three for the fifth grade.

In seeking to assess the consequential validity of the NTE, the present study attempts to answer the following research question:

*To what extent are the intended uses and consequences of the National Tests of English and their results evident in eighth-grade classrooms in Norway?*

The article first addresses the background of relevant previous research and validity theory, before examining the specific context of the NTE. The method of the two-part empirical study is then outlined before the results are presented and discussed in light of both validity theory and previous research. The paper concludes by summarising the findings and recommending further research.

This study has significance for the Norwegian context especially, as the NTE are taken across the whole country. The research presented here nevertheless has wider significance as it deals with the relationship between test validity and the actual implementation of test results in classrooms. As much of validity deals with intended uses and arguments for those uses, it is important to examine examples of *actual* test use in classroom settings (Moss, 2015). These examples serve to demonstrate the relationship between validity theory and what occurs in classrooms, thus contributing to a more complete picture.

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

**Previous research**

Given the relatively small size of Norway and its education system, there has been little research on the NTE and their uses. A study into the consequences of the eighth-grade English tests is therefore of value. Recent research on the use of the National Tests in Norway does exist, but it has not focused specifically on the English tests (Seland et al., 2013; Gunnulfsen, 2018; Roe et al., 2018; Gunnulfsen & Roe, 2018). Gunnulfsen (2018) conducted research on the ways in which the National Tests as a whole (Norwegian reading, mathematics, and English) were used in classrooms, concluding that teachers' relationships and attitudes towards the tests were largely influenced by the culture promoted by school leaders. Seland et al. (2013) found that the culture perpetuated by both teachers and school leaders was deemed critical to pupils' engagement with the tests and was largely shaped by teachers' and school leaders' views on the usefulness of the tests and their results. Similarly, Vestheim (2018) examined the practices around National Tests in schools with good results, finding that the results contributed to discussions around quality development in the school, and Mausethagen et al. (2017) examined the formats these discussions take. Mausethagen et al. (2019) used data from the PraDa project, which was a four-year investigation into the use of test data in Norwegian schools and municipalities. Among their findings was a broad agreement between school leaders and teachers about how National Test data should be used in classroom development, with school leaders being slightly more concerned with the importance of the results. Crucially for the present study, teachers generally thought that the National Tests can provide important information.

Sibbern (2013) completed a master thesis on the use of the English results and concluded that teachers and schools did not engage with the available materials and the results as much as intended by the Norwegian Directorate for Education and Training (*Utdanningsdirektoratet,* henceforth *Udir*). Lie et al. (2004) reached a similar conclusion with the now-discontinued tenth-grade English tests, and both noted that school leaders were largely responsible for schools' levels of engagement with the NTE. Although the studies mentioned here investigate the National Test results, both as a whole and the English tests specifically, none of them focus on the question of test validity in the way that the present study does.

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

Previous validation research on English reading tests has been wide-ranging, notably on the Test of English as a Foreign Language (TOEFL), which is unsurprising given its widespread international use (Educational Testing Service (ETS), 2011). The ETS summary of the validation research of the TOEFL tests includes details of research that has specifically focused on the uses and consequences of the test results and what this means for validity. This includes Wall and Horák's (2006, 2008, 2011) multi-year studies which specifically focus on one of the intended consequences of the TOEFL tests: to aid the teaching and learning of English. This is clearly a formative purpose, and also one of the stated purposes of the NTE. The ETS summary also includes a list of intended consequences/propositions against which validation studies are conducted, such as: "Performance on the test is related to other indicators or criteria of academic language proficiency" (ETS, 2011, p.5). The present study uses a similar list for the NTE, found in the 'history of the NTE' section.

**Theoretical background**

The overall theme of the study is the validity of the NTE. Messick (1989) defines validity as "the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions on the basis of test scores or other modes of assessment" (1989, p. 13). Messick acknowledges that decisions are always made on the back of test results; validity can then be argued to involve an assessment of these actions. He also identifies six main aspects of construct validity: content, substantive, structural, generalizability, external, and consequential aspects (1996). This study focuses primarily on the consequential aspect of validity. Consequential validity concerns the consequences of a test and how these consequences shape the context in which the test occurs.

Kane (2006) takes up Messick's concept of construct validity in his argument-based approach, later referred to as the measurement argument. He argues that the arguments for a test's validity can be divided into two: the interpretive argument and the validity argument. Kane describes interpretive arguments as setting out "the proposed uses and interpretations of test results by laying out a network of inferences and assumptions leading from observed performances to the conclusions and decisions based on the assessment scores" (Kane, 2006, p. 23). The validity argument then acts as an evaluation of these interpretations by examining the plausibility of their inferences and assumptions i.e., the extent to which they can be achieved.

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

Applying Kane's approach to the NTE, the interpretive argument for the validity of the NTE would focus largely on the uses and interpretations of the results, with the idea being that using the results as intended helps align the test with the curriculum and thus aid pupils' learning in a formative way (Pellegrino et al., 2016). In the case of the NTE, there is a clearly stated objective for the tests, as well as guidelines for schools, parents, and (especially) teachers as to how the results should be used and interpreted. These guidelines are codified in documents produced by *Udir*, as well as in the test administration system (PAS), which is a results analysis programme for teachers. These documents could be described as the outcome of constructing what Bachman and Palmer (2010) call an Assessment Use Argument (AUA). They describe an AUA as a set of statements that outline the proposed uses of a test, including score interpretation and consequences of results, something which "should be made explicit at the outset to guide development efforts and position a testing system to better achieve those intended inferences of consequences" (Chalhoub-Deville & O'Sullivan, 2020, p.152). Bachman and Palmer propose that an AUA should be dynamic and can change according to external conditions and requirements, even during an ongoing assessment. The NTE are largely shaped by external requirements, namely those of national educational authorities, necessitating a codified AUA reflecting these requirements.

Chalhoub-Deville (2015) asserts that Kane's argument-based approach could go further and argues that a social impact assessment can contribute to the integration of intended and actual consequences of a test, as well as more clearly defining the role of individual stakeholders (Chalhoub-Deville, 2009a, 2009b). Chalhoub-Deville points to the call from Bennett *et al.* (2011) to collect data from stakeholders with a view to not only establishing the intended consequences of a test, but also the *unintended* consequences, thus offering an insight into the social impact of a test. More recently, Chalhoub-Deville and O'Sullivan (2020) expand on the idea of consequences and stakeholders by drawing up a list of important considerations to be made when assessing the validity of a test based on its consequences. These considerations include which stakeholder groups to focus on and how these stakeholder groups perceive the consequences in question. The present study prioritises the stakeholder group of English teachers and their perceptions of the consequences of the NTE. In order however to establish what these consequences are, Moss (2015) argues that a potential disconnect must be overcome: that which lies between the intended uses and interpretations of

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

test results found throughout the field of validity theory and the *actual* uses and consequences in the classroom. Moss argues that consequences are invariably shaped by other factors, such as the students' learning before and after the tests, and the local capacity to put test results to good use. The following quote is directly applicable to the case of the NTE and the context which dictates their use:

> If the goal is to make decisions about how to improve teaching and learning or to make choices among alternative courses of action or policies, evidence of student outcomes alone is insufficient; one must consider information about the conceptual and material resources, the teaching processes and practices, and the organizational routines and cultures that shape or influence those outcomes. (Moss, 2015)

The next section of the article expands on these routines and cultures which surround the NTE, and which give rise to the need to examine the actual uses and interpretations of the results. The study itself explores the relationship between the intended uses of the tests and these routines and cultures; as Moss points out, these cannot truly be separated.

**The history and stated purposes and uses of the NTE**

The National Tests of English are made at the University of Bergen (UiB) on behalf of *Udir*. Since the proposal of the National Tests (English, Norwegian reading, and calculation) in 2002, as part of educational reforms in Norway (Hatch, 2013), there has been a degree of resistance from certain political parties, teachers, and pupil groups, who argued that the tests, for various reasons, do not achieve their stated aims (Regjeringen, 2002; Modal, 2005; NTB, 2005a; 2005b; 2005c, Carlsen, 2008). Upon conception, the general goal of the NTE was to:

> give schools an insight into pupils' fundamental skills in […] English. The information from the tests shall form a basis for formative assessment and quality development at all levels in the school system. (Directorate for Education and Training, 2019, author's translation).

The stated goal can be interpreted as the NTE being a tool for teachers to learn more about the competence of pupils with whom they are not familiar after their move from primary school to lower secondary school after the seventh grade. This would be supported by the first of the main intended consequences: forming a basis for formative assessment. Hattie and Timperley's (2007) framing of formative assessment inspired the development of formative assessment practices in Norway (Hasselgreen & Ørevik, 2020). It involves

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

feedback as a key component, which leads to further learning, empowering the learner themselves, with help from teachers, to assess where they are in their learning process, and where they are heading. A test that provides "an insight" into pupils' ability gives teachers something to feed back to pupils, providing at the very least a starting point from which to base further learning.

However, it is noticeable that the stated aim mentions providing "schools" with information, not just teachers. This broadens the goal somewhat from being an exclusively in-classroom tool to a wider measurement tool, involving multiple stakeholders. These stakeholders can be within the school itself and can also include stakeholders at a regional and national level. This is supported by the description of the tests' purpose from The Ministry of Education and Research (*Kunnskapsdepartementet*), who stated in a 2006 commissioning letter that:

> The National Tests will identify the extent to which pupils' skills are in accordance with the curriculum's aims for the basic skills in […] English, as they are integrated in the competency aims for subjects in LK06 after the 4th and 7th grades. The tests will provide information to students, teachers, school administrators, guardians, school owners, regional authorities, and national authorities as a basis for improvement and development work. (Udir, 2019, author's translation)

Thus, the NTE are intended to be used by both internal and external stakeholders, with varying levels of proximity to the NTE. Use by multiple stakeholders can be seen as a basis for the final part of the tests' stated goal: quality development at all levels in the school system. Based on the commissioning letter, this quality development is, at least partly, based on the information the stakeholders receive from the NTE.

The present study uses stated targets as exemplifications of intended consequences of the NTE, or an Assessment Use Argument. These intended consequences, and the extent to which they are met, are used by the study to assess the consequential validity of the NTE. Where these consequences are not met, the causes of them not being met are presented as the causes

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

of potential threats to validity. According to the stated goals and supporting documents (see next section), the intended consequences, not necessarily in any order, are as follows:

1. **Results forming a basis for formative assessment.**
2. **Results forming a basis for quality development.**
3. **Provide information to teachers and pupils about pupils' English skills through mastery level system and results breakdown.**
4. **Provide information to parents.**
5. **Provide information to regional and national authorities.**

**Teacher's guide**

The teacher's guide document (*lærerveiledning*), along with information on *Udir*'s website, give the clearest indication available to teachers of the intended *in-classroom* uses of the tests. The annually updated document is produced at the University of Bergen by the test developers and *Udir*. It is a comprehensive document that includes technical details, explanations of the tests' content, advice on preparation for the tests and, most crucially for this study, advice on how to interpret, and act on, test results. It is therefore intended to be a tool that aids the formative assessment and quality development mentioned in the tests' purpose.

The section on test results explains the five mastery levels that are used in the NTE for eighth grade and what they mean for a pupil's ability. This section also includes recommendations as to how the levels can be presented to internal and external stakeholders, which can contribute to quality development, and describes example scenarios as to how to raise pupils' mastery levels, contributing to formative assessment.

Another part of the teacher's guide is a breakdown of how the individual items from the NTE test specific reading processes, with suggestions as to how teaching can be adapted to enhance pupils' English reading skills. Given that the NTE are taken at the start of the school year, it is reasonable to assume that these suggestions for teaching and activities are intended as follow-ups to the test that teachers can incorporate based on the results. Indeed, *Udir* recommends that learning be adapted based on results throughout the school year, as opposed to in the immediate aftermath of the tests.

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

This study uses the teacher's guide as a reference point against which teachers' descriptions of how they use the test results can be compared. Based on this, validity can be at threat if there is a clear discrepancy between the recommendations in the teacher's guide and how teachers prepare for the test, interpret the results and how the results affect future teaching.

## Methodology

The present study includes a quantitative survey focusing on teachers' attitudes and uses of the NTE, followed by qualitative interviews seeking to establish the relationship between the intended and actual consequences of the NTE. It is therefore a mixed-method study with an exploratory sequential design (Johnson et al., 2007; Creswell & Clark, 2011) in that the qualitative interviews follow the quantitative survey and, to an extent, explore some of the data revealed by it. The research design is appropriate because it allows for multiple approaches to obtaining information about the use of the tests. A quantitative survey offers a perspective on the frequency of the recommended measures taken before and after the tests, as well as offering a window into attitudes towards the tests. The qualitative interviews offer a deeper view of teachers' thoughts and experiences surrounding the NTE and the associated routines, which can reveal a more nuanced picture of their views than a quantitative survey. Four core elements were deemed necessary to answer the research question and therefore constituted the backbone of the survey and interviews:

- How much emphasis the school places on the NTE
- How the teacher/department prepares for the NTE
- How the results are followed up
- Teachers' perceptions of the NTE

While the second of these categories, preparation for the NTE, does not directly concern the consequences of the results after the tests are taken, documents such as the teacher's guide offer advice for preparation, indicating it to be an intended consequence.

### Quantitative survey

The primarily quantitative digital survey was created using Survey Xact. It was sent to a distribution list of around 200 schools randomly selected by *Udir* each year to invite participation in piloting of the following year's NTE. Fully completed surveys from 37

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

schools were returned. The survey consisted of just six questions, primarily for reasons of brevity, to increase the likelihood of response (Jepson et al., 2005; Guo et al., 2016). The questions served a primary purpose of eliciting information about the frequency and levels of engagement with the NTE and the accompanying resources, with the additional purpose of 'setting the scene' for the qualitative interviews. The survey was written in Norwegian in the hope of maximising responses, assuming that most English teachers in Norway are Norwegian speaking, but results are presented in English in this article.

The first five questions were closed multiple-choice questions, two of which allowed respondents to select more than one answer, pertaining to individual practices that teachers might employ in the classroom. One of the four questions asked participants to rate the usefulness of the NTE on a Likert scale from 1 to 5. While an argument can be made that a five-point Likert scale can invite participants to 'settle' for the mid-range score, it was appropriate here as opinions on the usefulness of the NTE can be based on a multitude of positive and negative factors, meaning that not all scores were equally likely (Wagner, 2010).

The final question was an open question which yielded qualitative data: "Do you have any other comments about the use of the NTE in the classroom?". The purpose was to include comments on factors which the researcher may not have thought about to then be considered for inclusion in the interview guide if they were seemingly significant factors mentioned by multiple respondents.

Before distribution, a small-scale pilot of the survey was conducted at UiB, with the piloting participants having backgrounds in teaching English in Norway. The pilot did not raise any issues with the phrasing of the questions.

**Qualitative interviews**

The second part of the study involved semi-structured in-depth interviews with eighth-grade English teachers. The advantages of a semi-structured interview, as opposed to a standardised interview, are that they allow for probing and clarification of answers through follow-up questions (Kvale & Brinkmann, 2015), offering a richer data set (Holliday, 2010). In addition, the varying circumstances of the participants, such as the different ways schools utilise the NTE, preclude a fully standardised interview (Barriball & While, 1994). The

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

questions were largely open-ended, to encourage participants to reflect on their teaching practices and the NTE.

The six interviewees were selected through purposeful sampling (Palinkas et al., 2013; Miles & Huberman, 1994) and were English teachers who teach or have recently taught eighth-grade pupils and who have experience with the NTE. The participants were part of a larger group of teachers engaged by UiB as supervisors for students in teaching placements and were contacted by email. These teachers worked at different schools across six municipalities, which included schools in rural and urban areas, with pupils from a range of socioeconomic groups, thus allowing for a sample group that was broadly representative of Norway as a whole. Due to the COVID-19 pandemic, interviews took place remotely over the Zoom platform and were recorded with Zoom's recording tool for transcription. Interviewees were informed in the invitation email and immediately before the interview that their names, schools, and municipalities would remain anonymous. All interview recordings were deleted after transcription. Given the anonymous nature of the quantitative survey, it was not known if the interviewees had previously responded to that survey. The study was confirmed as being legal in relation to data protection laws by the Norwegian Centre for Research Data (NSD).

The analysis of the data was carried out using a form of theoretical thematic analysis (Braun & Clarke, 2006), whereby the data are analysed according to the theoretical framework, which in this case took the form of the four core elements. Data from the survey and interviews were extracted where directly relevant to the four elements and, after this, themes which emerged or were raised by multiple respondents. The results and discussion in the present article are presented and grouped according to this analysis. The results and discussion are presented together to stick to the theoretical framework, as opposed to the order of the interview and survey questions. Dissenting views between respondents are presented so as not to offer a false picture of unanimity.

**Research bias**

As researchers can be inherently biased (Baker & Gentry, 2006), any of the researcher's views needed to not affect the answers given by study respondents in any way. Qualitative interviews can be at risk of bias in the following ways: "the way we present ourselves and our study to respondents, […] the kind of questions we ask, and […] the way

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

we treat responses" (Weiss, 1994, p. 212). To address this, input from the researcher was limited, and questions were as open-ended and neutral as possible.

The researcher had no prior contact with the schools involved in the qualitative interviews and was unfamiliar with their current or previous test results or practices. Care was taken to explain to respondents that the researcher was not judging their teaching practices or their schools. This was especially important considering the researcher has previously worked with the development of the NTE; respondents were made aware of this and of the fact that the researcher was not studying on behalf of the test developers or *Udir*. The names of the schools that responded to the digital survey were not available to the researcher.

## Results and discussion

In this section, the results and discussion are integrated, organised according to the core elements identified in the methodology section. They are not divided by the individual intended consequences detailed in the 'history of the NTE' section, as many answers overlapped and could apply to multiple consequences. After the four core elements are presented, significant themes that emerged from the interviews outside of the core elements are discussed. The full results of the survey can be viewed as appendices.

### Emphasis placed on the NTE

The first of the core themes aimed to 'set the scene' in that it sought to elicit descriptions of the practices surrounding the NTE in schools, thus revealing prevailing attitudes and degrees of interaction with the tests. The general emphasis that schools place on the NTE was also viewed as an indication of attainment of the key intended uses of the tests: as a basis for formative assessment and quality development.

As this was a largely open theme, in that descriptions of schools' emphasis on the tests were sought, the interviews provided a richer data set for the theme. The answers covered both extremes: the school placing no emphasis at all on the tests, with no preparation or follow-up, and schools placing a great deal of emphasis on the results of the NTE, with follow-up meetings arranged to discuss what the results mean. Respondents 1 and 3 for example described a general lack of engagement with the tests in recent years, describing them as "a formality" (respondent 1). Conversely, respondent 2 reported that their school had paid little attention to the tests previously but that, during the last three years, "we've kind of

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

looked at the results and seen who is doing well and who's not doing well and tried to present some extra effort on those who are struggling".

A way in which schools' emphasis on the tests manifested itself was described by respondents 1, 3 and 6, namely school principals placing an unofficial status on the tests and their results, based to a large extent on comparison with other schools' results. Respondents 1 and 3 went as far as to mention pressure being previously placed on teachers to improve results in relation to other schools. This pressure and comparison between schools can be described as an unintended consequence (Chalhoub-Deville, 2015) as it is not mentioned in the official purpose or the teacher's guide as being part of the intended uses of the NTE. However, the fact that teachers (and principals) can see how their pupils performed in relation to the rest of the country through the PAS system, means that some degree of comparison is inevitable; it is how this comparison is responded to that decides whether the intended uses of the NTE are met. As noted by some of the survey respondents in question 6, the fact that the tests are taken so early in the school year means that to expect eighth-grade teachers to improve results is unfair and would not meet the tests' purpose as a formative tool. However, if results show a school to be underperforming in relation to others, and decisions are made by principals to remedy this, there is a clear argument that the intended use of the tests as a basis for quality development is evident. This would support the interview respondents' views that the emphasis placed on the NTE is largely dependent on principals' perspectives on their significance.

**Preparation for the NTE**

Possibly because the NTE are taken shortly after the beginning of the school year, leaving teachers with little time, respondents generally reported a lack of preparation for the tests outside of technical preparation. The most significant difference among the interviewees was the level of awareness of the resources available to them to aid preparation. There appeared to be an awareness of the availability of previous tests and example tests on *Udir*'s website, with 73% and 59% of survey respondents reporting utilising the respective resources. However, only 32% reported consulting the teacher's guide as part of preparation, with both survey respondents and interviewees reporting a lack of awareness that the teacher's guide

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

could be used for preparation. Respondent 5 expressed regret about not preparing more for the tests outside of technical preparation for the format of the tests:

> "It's kind of one of the things afterwards when you read the lærerveiledning (teacher's guide), it's ahhh, I should maybe have read this before I did the tests, but I don't spend much time before."

The lack of engagement with the teacher's guide (and the recommendations within it) was also reflected in several comments in the survey which reported either a total lack of awareness that the NTE could be prepared for, or a lack of willingness to engage with the tests:

> "It happens too early in the school year. We have other things we'd rather be doing, like getting to know the pupils and creating a good and safe environment in the class, as opposed to spending time on the National Tests. They always seem to be 'thrown' down onto us."

The above comment also alluded to the timing of the tests being an issue, a theme that is discussed later in this section.

**Follow-up of results**

The third key theme of the study focused on interaction with the results and is arguably the core theme which is most decisive in whether the intended consequences of the NTE can be said to be present. As was the case with the general emphasis on the NTE, the results offered a picture of a lack of uniformity. The second question of the digital survey asked if teachers gave detailed feedback to pupils on an individual basis, a whole class basis, both, or neither. The most popular answer, with 46%, was 'neither', which suggests that the formative purpose of the tests is possibly *not being met*, based on Hattie and Timperley's (2015) characterisation of feedback being essential to formative assessment. The result could also be looked at as 54% of respondents reporting that results *are* fed back to pupils in some way, indicating the use of the tests as part of a formative assessment process. The truth may well lie somewhere in the middle, as the split is close to 50-50. The lack of feedback in almost half of the cases can indicate a *potential* threat to validity, based on Messick's (1989) and Kane's (2006) ideas about the strength of the arguments used to support the results of the test being utilised as intended. The

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

idea of *detailed* feedback may have had a slight impact on the result; some teachers may give brief feedback, but not enough that they consider it detailed.

Feedback to pupils is of course not the only way in which the results can be followed up, and this was reflected in the interviews. Respondents 2 and 5 for example described comprehensive follow-up, with meetings arranged to discuss results attended by teachers, department heads, administrators, and special education coordinators. The respondents noted that special attention was paid to pupils whose results placed them below mastery level three, with decisions being made as to whether they required extra teaching to improve their skills. This process could be argued to fulfil both main purposes of the tests – using results as a platform to adapt teaching to under-achieving pupils has a clear formative purpose for the pupils in question, and adapting teaching generally based on the results can be said to contribute to a process of quality development. This is supported by the fact that respondent 5 reported English teachers at the school being sent to courses specifically focused on the use of the NTE and their results. This clear attempt to improve the processes around the tests and the results can certainly be described as part of a quality development process.

The interviews did not entirely offer a picture of comprehensive interaction with the results. Respondents 1, 3 and 4 described a declining level of significance in terms of how their respective schools viewed the tests. All three said that the results were consulted, but only to check if there were any alarming changes (respondent 4) or to check if any pupils achieved below mastery level three (respondent 3), while respondent 1 reported individual teachers examining results to become familiar with pupils' abilities but no organised meetings or follow-up. Despite the respondents presenting their admittedly brief consultations with the results as being symbolic of the tests' lack of significance in their schools, it does not necessarily mean that the tests are not being used at least partly as intended. All three of these respondents described using the tests as a source of information, which corresponds with the goal of "giv[ing] schools an insight into pupils' fundamental skills in […] English" (Udir, 2019). This insight, despite not being explicitly acted upon and therefore not used as a basis for formative assessment or quality development, presents an argument for consequential validity, albeit a weak one, because teachers are at least provided with information.

Both the digital survey and the interviews presented evidence of teachers using the results as sources of information, to varying extents, which would appear to suggest that the

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

NTE results are interacted with as intended. However, in describing the results in the form of low scores as a 'warning signal', indicating which pupils fall below certain mastery levels and, in the case of schools that actively follow up results, require extra teaching or attention, part of the tests' purpose arguably falls by the wayside. The stated purpose of the tests refers to "pupils", as opposed to merely pupils for whom the English subject is proving difficult. The Ministry of Education and Research's commissioning letter (Udir, 2019) refers to the curriculum and the quality development aspects of it, which also includes the right to adapted education for all pupils, regardless of ability level. There is therefore a strong argument that, by focusing on only the pupils who fall below a level of concern, the pupils whose scores place them in the higher ability range, and who are equally entitled to teaching adapted to their level, are somewhat forgotten in relation to the NTE. This suggests that a significant portion of the information the tests provide is not acted upon, thus presenting a potential threat to validity.

**Teachers' perceptions of the NTE**

This core element focused on teachers' perceptions of the tests themselves and their perceptions of the stated purposes. The fifth survey question asked respondents to rate how useful they perceived the tests to be on a five-point Likert scale, with 1 being least useful and 5 being most. The mean score was 2.64, with a median of 3. The more negative scores, namely 1 and 2, were chosen by 17 respondents, or 46%, as opposed to only seven respondents, or 19%, choosing scores of 4 and 5. This negatively weighted response may reflect some of the general feelings towards National Tests as a whole, as opposed to the content of the National Tests of English specifically, reflected by some of the comments in the sixth survey question:

> "I think there is a large distance between the content of the National Tests and the content of the curriculum. There are too many National Tests in general, and I don't really see the value in the time spent on them or in the unnecessary pressure they place on pupils"
>
> "Pupils become demotivated by tests such as these; it's good that I can see what they struggle with, but the level is so much higher than that which my pupils can perform at that the negative consequences are more noticeable than the positive. No to mandatory National Tests."

As alluded to in the section on the history of the NTE, there was a degree of resistance from some groups, including some teachers, to the National Tests as a whole, and there appear

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

to be elements of this resistance remaining in some of the survey responses. It should be noted that, during the interviews, perceptions of the tests were much more nuanced, even among those who reported a lack of engagement with the tests at their schools. This is not to say that perceptions of the tests and their purposes were overwhelmingly positive among interviewees; respondents 5 and 6 did not feel that the tests could realistically serve as a reliable source of information as they were taken on one day, so they felt that performance could be affected by other factors. Instead, they believed that continuous assessment throughout the year would offer a more reliable picture. It must be acknowledged here, however, that the tests are intended to be a tool which can *contribute* to formative assessment, as opposed to being its sole component. One could argue here that this limitation is not communicated effectively enough in the tests' stated purposes, as respondents 1, 2 and 3 also viewed the stated purpose as "lofty" (respondent 1), maintaining that the results could only contribute to formative assessment and quality development, as opposed to forming a basis for them. The teacher's guide document makes suggestions as to how the results can be integrated into a programme of formative assessment but, as noted by multiple interviewees, studying this requires time not always available to teachers:

> "And I think it would have been incredibly nice to have such a good tool as the National Tests actually are and to be able to use it for what it's actually meant for, in a good way, in a positive way, but it requires that something else has to give way for us to be able to focus on it". (Respondent 3)

This quote exemplifies the idea expressed by some respondents of the NTE actively competing for time with other duties and activities, with occasionally negative consequences in the form of increased time pressure felt by teachers, which will be expanded upon in the next section.

**Time pressures and timing of tests**

One of the most prominent themes to emerge from the survey and the interviews was that of time, specifically the time afforded to the NTE and the results and the time during which the tests are taken.

A lack of time afforded to the tests was cited by all the interviewees in one way or another, whether that be time to analyse the results or time to consult other information sources such as the teacher's guide. Despite some of the comments made in the survey, there appears

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

to be a willingness to engage with the NTE and a recognition of the potential benefits on the part of teachers. However, it is not something all teachers can prioritise, especially those for whom support is not forthcoming from the principal level of the school. It is probably no coincidence that the interviewees who most viewed the information provided by the tests as useful, respondents 2 and 5, were those who were afforded the most time to follow up on the results in the form of meetings, discussions with parents, and feedback to pupils. However, even these respondents noted that they would benefit from using the teacher's guide, but that time prevented them from doing so. These results, therefore, appear to show more of an acknowledgement on teachers' behalf of the *potential* usefulness of the tests, compared to, for example, the conclusions reached by Seland et al. (2013), although the present study has an admittedly narrower focus.

In terms of wider implications, time pressure is a threat to validity that can be argued to be somewhat neglected in validity research, especially considering Moss's (2015) discussion of the 'real life' consequences in classrooms as opposed to the decisions made at other stakeholder levels. Much of validity theory presents test use as something of a conscious choice, such as Bachman and Palmer's (2010) AUA, whereas the results of the present study offer a picture of teachers who wish to engage more with the results of the NTE but who are unable to do so, primarily due to time constraints. Of course, one can also argue that the lack of time afforded to following up the results in schools is in itself a conscious decision and merely a sign of the general culture around the National Tests in individual schools, as opposed to a reflection of views on the National Tests of English. This is supported by the respondents who reported a lack of engagement with the results saying it was the case for all National Tests, not just the eighth-grade English tests.

Some survey respondents mentioned the time that the tests are taken (the beginning of the school year) as being an issue. They stated that they are taken "too early in the school year" to offer any information of value, especially given the fact that pupils in the eighth grade are new to lower secondary school so teachers are not familiar with them, nor with their English reading skills.

As the timing of the tests was a theme that emerged from the digital survey, it was investigated further during the interviews, and interviewees were asked what impact, if any, they felt that taking the tests later in the school year would have. Respondents 3 and 4 indicated

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

that the results would be different, as the teachers would have a full year to work with the pupils, but that they would be much less useful, as the tests would lose their purpose of providing information about pupils with whom teachers are unfamiliar. Respondent 1 alluded to pressure being applied to teachers to deliver good results and believed that better results could be achieved if the tests were taken later in the year, when teachers are more familiar with the pupils' strengths and weaknesses. However, the respondent acknowledged that this pressure has been steadily decreasing and that the tests can be very useful as they are, if the results are not subject to intense media scrutiny, which the respondent felt was a key source of pressure on teachers. Concern from teachers with regard to pressure being placed on them to achieve better results in the NTE is certainly understandable and very much representative of an unintended consequence and can thus lead to a threat to validity (Chalhoub-Deville, 2015). However, as interviewees acknowledged, the formative aspect of the results would be lost if the tests were taken later in the year, and there is no guarantee this would alleviate pressure to deliver better results; the opposite may likely be true, given the often higher-stakes nature of year-end summative assessments. It is also possible to reflect on the ways in which this pressure manifests itself; it may not always be negative, especially in terms of quality development. For example, respondent 5's description of teachers being sent on courses about the NTE included "the fact that I knew I was going to participate in a course made me work through my results from my class". This pressure felt by the teacher, and the content of the course itself, clearly impacted engagement with the results and thus had potential implications for quality development and formative assessment practices.

**Parents and external stakeholders**

Despite the focus of the present study being teachers' uses of the NTE and what they mean for validity, parents were mentioned by some of the interviewees. Respondents 2 and 4 for example described the results of the NTE as being key components of conversations with parents, describing the mastery levels used by the NTE as means of framing and quantifying their child's ability in the English subject. Conversely, respondents 3 and 6 reported almost non-existent relationships between parents and the NTE, with parents very rarely enquiring about results.

How much parents' engagement with test results affects validity can depend on how much one chooses to prioritise parents as stakeholders, based on Chalhoub-Deville and

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

O'Sullivan's (2020) list of considerations. The commissioning letter (Udir, 2019) lists parents as stakeholders for whom the NTE should provide information, along with administrators, school owners, and local and national authorities. Therefore, interviewees reporting interaction with parents suggests the presence of intended consequences. This is supported by the fact that none of the interviewees said that a parent had ever asked how their child's school performed compared to other schools; their interest has only been in information about their own child's ability. However, as demonstrated by respondents 3 and 6, this interaction with parents is not universal. If parents are not engaged with pupils' results, the (external) stakeholders mentioned in the tests' official purpose become limited to local, regional, and national authorities. This can be seen to invite comparisons of results between schools, and with it pressure on teachers, one of the key reasons given for the objections to the National Tests upon their introduction to Norway (Carlsen, 2008). This serves to further demonstrate Moss's (2015) highlighting of the potential disconnect between intended uses of tests and results and the actual carrying out of these uses in schools.

**Overall assessment of validity**

To assess the (consequential) validity of the NTE, the tests must be placed in the context of the assumptions made about the consequences (Kane, 2006). If the intended consequences listed in the 'history of the NTE' section are interpreted as assumptions, the validity argument then assesses the extent to which these intended consequences can be argued to be present. The most obvious, and significant, obstacle to these intended consequences being achieved is a lack of engagement with the results. 46% of survey respondents reported not providing any feedback to pupils at all, and four of the six interview respondents reported either no follow-up of the results or follow-up being sporadic, when the schools face alarming changes in performance. This lack of engagement would suggest that the results of the NTE, in a little under half of the cases in the whole study, are *not* being used as intended. This is the clearest *unintended* consequence found in the study, thus leading to the clearest threat to validity. However, the extent of the threat is somewhat harder to gauge.

Despite the lack of feedback given to pupils, all interviewees reported *some* degree of conversations between teachers about the results, even if in an unofficial capacity, with these conversations identifying pupils who require extra help. The main difference between respondents was the amount of *official* follow up to the results; this seems to be entirely

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

dependent on the internal culture of individual schools, corresponding with the findings of Seland et al. (2013) and Vestheim (2018). Assessing the consequential validity of the NTE then depends partly on how one interprets *Udir*'s stated purpose (Udir, 2019). If one understands the chief purpose of the tests to be a mere source of information for teachers, then most cases suggest the tests are being used as intended, whether officially or otherwise. However, if the results are intended to be an *official* basis for formative assessment and quality development, with results being actively engaged with as part of a structured follow-up programme, which is indeed recommended (but not mandated) by *Udir*, the results of this study can suggest that the aim is not being achieved in a considerable number of Norwegian schools.

## Conclusion and further study

Through interviews and an electronic survey, the current study establishes a clear lack of uniformity in the way in which the NTE are integrated into the formative assessment process for eighth-grade pupils. However, the sample size is somewhat limited, and a wider sample might have offered a more complete picture. The present study nevertheless offers a picture of teachers' current views of the NTE and their consequences, which is key for any assessment of consequential validity, especially in-classroom low-stakes tests such as the NTE.

While some schools actively prepare for, and act upon, the tests, utilising the resources made available by *Udir*, a lack of engagement, trust or both towards the tests causes some teachers to view them as an irrelevance or, worse, an active consumer of valuable time. The idea of the NTE being an unwelcome consumer of time seems to correspond with a lack of engagement with resources such as the teacher's guide; a renewed effort to promote the guide as a resource may offer a means of increasing engagement with the test results.

Any lack of engagement with the NTE appears to be a decision taken at a principal level, as opposed to an individual teacher level, supported by the fact that even the interviewed teachers who did not see the tests as useful resources said that they could see *potential* uses for them, albeit uses that they do not have time for. Given the reports from some teachers of pressure being applied by school principals to compete with other schools, and of other principals nurturing a culture of active engagement with the NTE results, it seems that principals have a significant influence on the presence of intended consequences.

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

A future study would therefore benefit from interviews with school principals, to establish what they base their decisions surrounding the NTE on, and if they create a culture of using the tests for purposes of formative assessment and wider quality development. Based on this study, principals appear to be the stakeholders with the most power to address potential threats to the consequential validity of the NTE.

**References**

Bachman, L.F., & Palmer, A.S. (2010). *Language assessment in practice*. Oxford University Press.

Barriball, K.L. & While, A. (1994). Collecting data using a semi-structured interview: a discussion paper. *Journal of advanced nursing 19*, 328-335. https://doi.org/10.1111/j.1365-2648.1994.tb01088.x

Baker, S.M. & Gentry, J.W. (2006). Framing the research and avoiding harm. In R. W. Belk. (Ed.) *Handbook of qualitative research methods in marketing* (p. 322-335). Edward Elgar.

Bennett R.E., Kane M.T., & Bridgeman, B. (2011, Feb 10-11). *Theory of action and validity argument in the context of through-course summative assessment.* [Conference presentation]. Invitational Research Symposium on Through Course Summative Assessment, Atlanta. https://www.ets.org/research/policy_research_reports/publications/paper/2011/imxq

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Carlsen, C.H. (2008). Crossing the bridge from the other side: the impact of society on testing. In L. Taylor & C.J. Weir (Eds.) *Studies in language testing 31: Language testing matters* (p. 344-356). Cambridge University Press.

Chalhoub-Deville, M. (2009a). The intersection of test impact, validation, and educational reform policy. *Annual Review of Applied Linguistics 29*, 118–131. https://doi.org/10.1017/S0267190509090102

Chalhoub-Deville, M. (2009b). Standards-based assessment in the U.S.: Social and educational impact. In L.L. Taylor & C.J. Weir (Eds.), *Studies in language testing 31: Language testing matters* (p. 281-300). Cambridge University Press.

Chalhoub-Deville, M. (2015). Validity theory: Reform policies, accountability testing, and consequences. *Language testing 33 (4)*, 453-472. https://doi.org/10.1177/0265532215593312

Chalhoub-Deville, M. & O'Sullivan, B. (2020). *Validity: Theoretical development and integrated arguments*. Equinox Publishing Ltd.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). SAGE Publications.

Educational Testing Service. (2011). Validity evidence supporting the interpretation and use of TOEFL iBT scores. *TOEFL Research Insight Series Volume 4*. Retrieved from https://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v4.pdf

Gunnulfsen, A.E. (2018). *Micro policy making in schools. Use of National Test results in a Norwegian context.* [Doctoral dissertation]. University of Oslo.

Gunnulfsen, A.E. & Roe, A. (2018). Investigating teachers' and school principals' enactments of national testing policies. A Norwegian study. *Journal of Educational Administration 56* (3), 332-349. Doi: 10.1108/JEA-04-2017-0035

Guo, Y., Kopec, J.A., Cibere, J., Li, L.C., & Goldsmith, C.H. (2016). Population survey features and response rates: A randomized experiment. *American Journal of Public Health 106* (8), 1422-1426. https://pubmed.ncbi.nlm.nih.gov/27196650/

Hasselgreen, A. & Ørevik, S. (2020). Assessment and testing. In A. Fenner & A.S. Skulstad (Eds.), *Teaching English in the 21st century* (p. 365-388). Fagbokforlaget.

Hatch, T. (2013). Beneath the surface of accountability: Answerability, responsibility and capacity-building in recent education reforms in Norway. *Journal of Educational Change 14* (2), 113-138. https://doi.org/10.1007/s10833-012-9206-1

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Holliday, A. (2010). Analysing qualitative data. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (p. 98-111). Continuum International Publishing Group.

Im, G.H., Shin, D. & Cheng, L. Critical review of validation models and practices in language testing: their limitations and future directions for validation research. *Language Testing in Asia 9* (14). https://doi.org/10.1186/s40468-019-0089-4

Jepson, C., Asch, D.A., Hershey, J.C., & Ubel, P.A. (2005). In a mailed physician survey, questionnaire length had a threshold effect on response rate. *Journal of Clinical Epidemiology 58 (1)*, 103-105. https://pubmed.ncbi.nlm.nih.gov/15649678/

Johnson, R. B., Onwuegbuzie, A.J., & Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research 1*, 112–133. https://doi.org/10.1177/1558689806298224

Kane, M.T. (2006). Validation. In R. Brennan. (Ed.), *Educational measurement* (4th ed.), (p. 17–64). American Council on Education and Praeger.

Kvale, S. & Brinkmann, S. (2015). *Det kvalitative forskningsintervju* (3rd ed.). Gyldendal Akademisk.

Lie, S., Caspersen, M. & Björnsson, J. (2004). *Nasjonale prøver på prøve*. Retrieved from https://www.udir.no/globalassets/upload/forskning/5/nasjonale_prover_pa_prove.pdf

Mausethagen, S., Prøitz, T., & Skedsmo, G. (2017). Teachers' use of knowledge sources in 'result meetings': thin data and thick data use. *Teachers and Teaching 24* (1), 37-49. https://doi.org/10.1080/13540602.2017.1379986

Mausethagen, S., Skedsmo, G, & Prøitz, T. (2019). Hva slags utvikling? Elevresultater som utgangspunkt for utviklingsarbeid. In K. Helstad & S. Mausethagen (Eds.), *Nye lærer- og lederroller i skolen* (p. 53-69). Universitetsforlaget.

Messick, S. (1989). Validity. In R.L. Linn. (Ed.). *Educational measurement*, (p. 13–103). American Council on Education and Macmillan.

Messick, S. (1996). Validity and washback in language testing. *Language Testing 13*(3), 241-256. https://doi.org/10.1177/026553229601300302

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). SAGE Publications.

Modal, E. (2005, January 18). 'Varsler boikott av nasjonale prøver'. *Nettavisen*. https://www.nettavisen.no/nyheter/innenriks/--varsler-boikott-av-nasjonale-prover/331356.html

Moss, P.A. (2015). Shifting the focus of validity for test use. *Assessment in Education: Principles, Policy and Practice 23*(2), 236-251. https://doi.org/10.1080/0969594X.2015.1072085

NTB. (2005, January 27). 'Omfattende boikott av nasjonale prøver'. *Stavanger Aftenblad*. https://www.aftenbladet.no/innenriks/i/1W9gG/omfattende-boikott-av-nasjonale-prver

NTB. (2005, March 28). 'Elevaksjonistene legger ut alle vårens prøver'. *Stavanger Aftenblad*. https://www.aftenbladet.no/innenriks/i/8OkEQ/elevaksjonistene-legger-ut-alle-varens-prver

NTB. (2005, April 3). 'Elevaksjonistene fikk ikke tak i engelskprøve'. *VG*. https://www.vg.no/nyheter/innenriks/i/nkGpd/elevaksjonistene-fikk-ikke-tak-i-engelskproeve

Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N. & Hoagwood, K.A. (2013). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research 42* (5), 533-544. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4012002/

Pellegrino, J.W., DiBello, L.V., Goldman, S.R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist 51*(1), 59-81. https://doi.org/10.1080/00461520.2016.1145550

Regjering. (2002). *Innstilling fra kirke-, utdannings- og forskningskomiteen om bevilgninger på statsbudsjettet for 2003 vedkommende Utdannings- og forskningsdepartementet, Kultur- og kirkedepartementet, Nærings- og handelsdepartementet, Fiskeridepartementet og Landbruksdepartementet.* (Budsjett-innst. S. nr. 12 2002-2003). Retrieved from https://www.stortinget.no/no/Saker-og-publikasjoner/Publikasjoner/Innstillinger/Budsjett/2002-2003/innb-200203-012/?lvl=0#a4.2.1

Roe, A., Andresen Ryen, J. & Weyergang, C. (2018). *God leseopplæring med nasjonale prøver. Om elevers leseutfordringer i et mangfold av tekster.* Fagbokforlaget.

Seland, I., Vibe, N., & Hovdhaugen, E. (2013). *Evaluering av nasjonale prøver som system* (Utdanningsdirektoratet report 4/2013). Retrieved from https://www.udir.no/globalassets/filer/tall-og-forskning/rapporter/2013/nifu_np.pdf

Sibbern, M. (2013). *The National Test in English: Why it is important and why it is not enough*. [Master's thesis]. University of Oslo.

NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

Utdanningsdirektoratet (2019) *Om nasjonale prøver*. Norwegian Directorate for Education and Training. https://www.udir.no/eksamen-og-prover/prover/nasjonale-prover/om-nasjonale-prover

Utdanningsdirektoratet (2019) *Hva er nasjonale prøver?* Norwegian Directorate for Education and Training. https://www.udir.no/eksamen-og-prover/prover/rammeverk-for-nasjonale-prover/hva-er-nasjonale-prover/

Vestheim, O.P. (2018). Nasjonale prøver – hemmende styringsverktøy eller lokale redskap for praksisutvikling? *Acta Didactica Norge 12* (4), Article 3. https://doi.org/10.5617/adno.6249

Wagner, E. (2010). Survey research. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (p. 22-38). Continuum International Publishing Group.

Wall, D., & Horák, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 1, the baseline study* (TOEFL Monograph No. 34). Educational Testing Service.

Wall, D., & Horák, T. (2008). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 2, coping with change* (TOEFL iBT Research Report No. 05). Educational Testing Service.

Wall, D., & Horák, T. (2011). *The impact of changes in the TOEFL exam on teaching in a sample of countries in Europe: Phase 3, the role of the coursebook. Phase 4, describing change* (TOEFL iBT Research Report No. 17). Educational Testing Service.

Weiss, R.S. (1994). *Learning from strangers: the art and method of qualitative interview studies*. The Free Press.

Nordic Journal of Language Teaching and Learning

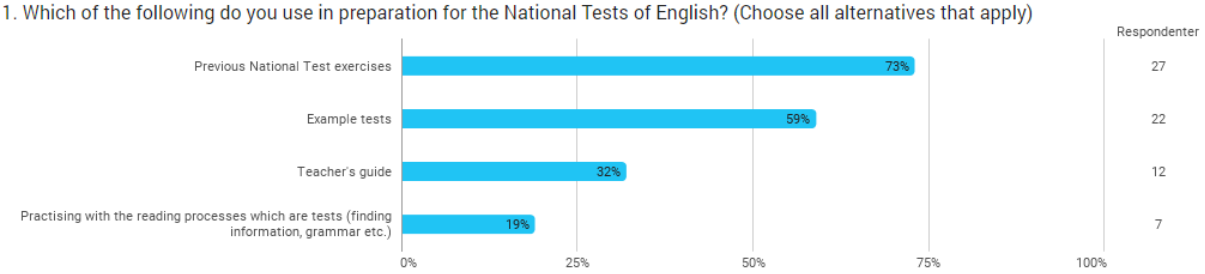
NJLTL Vol. 10 No. 1(2022) DOI 10.46364/njltl.v10i1.1017

# Appendices

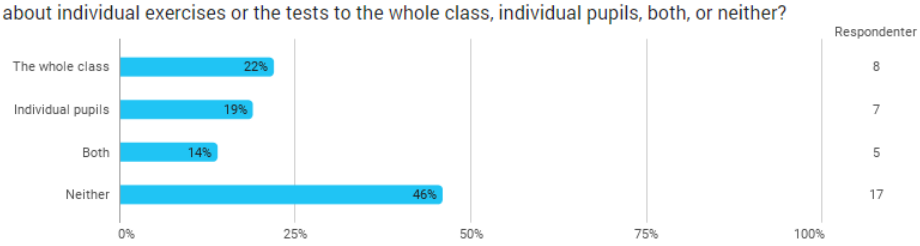
## Appendix A

## Survey Results

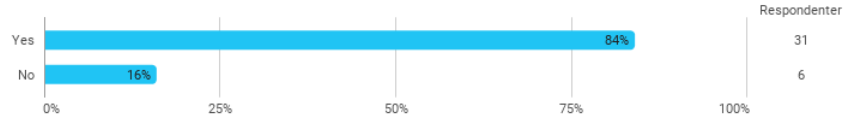
1. Which of the following do you use in preparation for the National Tests of English? (Choose all alternatives that apply)



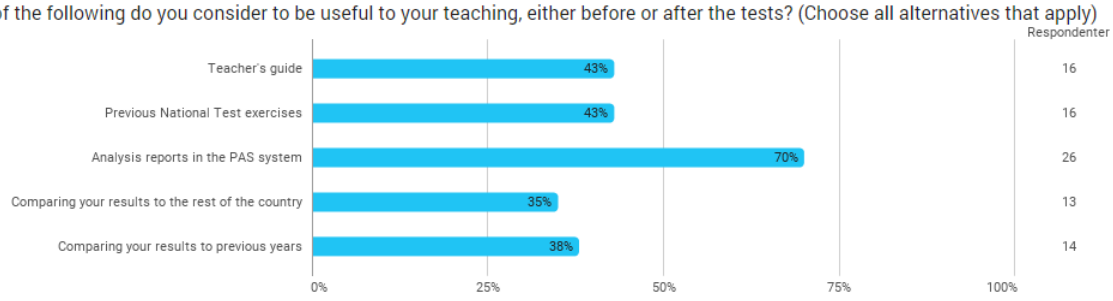

2. Do you give detailed feedback about individual exercises or the tests to the whole class, individual pupils, both, or neither?




3. Do you focus your teaching on the reading processes (finding information, understanding main points etc.) which the results suggest your pupils have difficulty with?




4. Which of the following do you consider to be useful to your teaching, either before or after the tests? (Choose all alternatives that apply)




5. How useful do you consider the National Tests of English to be to your teaching? 5 is very much so, 1 is not at all.