

A MAGABIZTOSSÁG ÉS A TELJESÍTMÉNYBECSLÉS PONTOSSÁGÁNAK KUTATÁSA AZ ANGOL GCSE VIZSGÁK ÉRTÉKELŐINÉL

Nádas Rita és Irenka Suto

Cambridge Assessment (University of Cambridge Local Examinations Syndicate)

Az angol GCSE vizsgák rendszere és értékelése

Az angol tanulók – a magyar fiatalokhoz hasonlóan – 16 éves korukig iskolakötelezettek. Ekkor tehetik le az úgynevezett „General Certificates of Secondary Education” (GCSE) vizsgát, amit a „Középiskolai Oktatás Általános Oklevelei”-nek lehetne fordítani. Angliában három nagy értékelő bizottság (*Awarding Body*) működik, ezek egyike a Cambridge Assessment, amely minden évben számos tantárgyban meghirdeti a GCSE vizsgákat. A tanulók átlagosan hét – nyolc választott tantárgyból tesznek GCSE vizsgákat (*Vidal Rodeiro, 2007*).

Nemcsak a választható tantárgyak sokfélék: egy adott tantárgyon belül is változatosak a vizsgafeladatok, a többszörös választástól a hosszú esszéig mindenféle gyakorlatot tartalmaznak. Ezeket a vizsgákat nem a diákok saját középiskolai tanára javítja, hanem külön erre képzett, a tananyagot szintén oktató megbízott szakértő értékelők.

A GCSE vizsgák eredményei igen fontosak: többek között ezeket az eredményeket veszik figyelembe az egyetemek, főiskolák, valamint a munkáltatók felvételi döntésében is szerepet játszhatnak. Éppen emiatt nagyon fontos, hogy a válaszokat pontosan és megbízhatóan javítsák az értékelők. A központok ezért szigorúan ellenőrzik a javítási folyamatot, és minden egyes értékelési időszak előtt speciális képzésben részesítik értékelőiket. Ezeket a tréningeket sztenderdizálásnak, vagy koordinálásnak nevezik (*Qualifications and Curriculum Authority, 2006*).

Greatorex, Nádas, Suto és Bell (2007) szerint a GCSE vizsgák éppen átalakulóban vannak. A legfrissebb kutatások az eltérő szakmai (oktatói- és értékelői) háttérrel rendelkező értékelők alkalmazására irányulnak (pl. *Suto és Nádas, 2007*). Erre azért került sor, mivel Nagy-Britanniában az utóbbi években néhány tantárgyban megcsappant a nagy szakmai tapasztalattal rendelkező értékelők száma, főleg a tapasztalt oktatók körében. Ezen kívül egyre nagyobb az igény a sokféle szakmai háttérrel rendelkező értékelők hatékony és optimális képzésének kialakítására.

A magabiztosság és a teljesítménybecslés pontosságának kutatása

Terjedelmes szakirodalom vizsgálja a magabiztosság és a valós teljesítmény pontos becslésének kérdéskörét a főiskolai hallgatók vizsga utáni önértékelésétől (pl. *Maki*, 1998; *Koch*, 2001) a munkához kapcsolódó önértékelésekig (*Dunning*, *Heath* és *Suls*, 2004). Azonban a GCSE értékelők saját teljesítmény-értékelését és metakognícióját – tudomásunk szerint – még nem vizsgálták.

A tesztértékelők ön-reflexiója több szempontból is lényeges. Először, amennyiben saját teljesítményük becslése helyesnek bizonyul, az értékelőtestületek (*Awarding Bodies*) ezt az információt fel tudják használni a sztenderdizációs folyamat¹ során arra, hogy körvonalazzák és megvizsgálják azokat a vizsga-feladatokat, amelyekkel az értékelőknek nehézségei akadtak. Viszont amennyiben az értékelők meglátásai megbízhatatlannak, és nem függenek szorosan össze értékelésük pontosságával, a „problematikus területekkel” kapcsolatos visszajelzéseik félrevezetőek lehetnek. Például, azon kaphatják magukat a vezető értékelők (*Principal Examiners*), hogy a sztenderdizációs folyamatban nem a megfelelő kérdésekre összpontosítanak. Másodszor, annak vizsgálata, hogy a magabiztosság és a teljesítmény becslésének pontossága változik, esetleg pontosabbá válik-e az értékelési tapasztalatok, illetve a visszajelzések számának növekedésével, kihathat az értékelő testületek értékelő-tréning során használt gyakorlatára, ami által pontosabbá válhat az értékelés: a kutatások azt mutatják, hogy a saját teljesítmény becslési pontosságának fejlődése jobb teljesítményt eredményez (*Koch*, 2001; *Dunning*, *Johnson*, *Ehrlinger* és *Kruger*, 2003).

Jelen tanulmányban bemutatjuk azon kutatás céljait és eredményeit, amely feltárja a GCSE értékelők reflexióit saját értékelői teljesítményükkel, nevezetesen értékelésük pontosságával kapcsolatban. Vizsgálatunk az értékelők magabiztosságának szintjére és teljesítménybecslésük pontosságának szintjére, valamint ezek esetleges változásaira terjedt ki az értékelési eljárás során. A „magabiztosság” (*self-confidence*) terminus itt arra vonatkozik, miként becsülik az értékelők az eljárás után, mennyire pontosan értékelték az adott kérdéssort; a „rálátás”, (*insight*), „kalibráció”, (*calibration*, *realism* lásd *Stankov* és *Kleitman* 2008), pedig a valós és a becsült pontosság közötti kapcsolatot jelöli, megmutatva a becslés precizitását.

A teljesítménybecslés kalibrációja és a magabiztosság elméleti keretei

A teljesítménybecslés kalibrációját széles körben, többféle perspektívából vizsgálták már a kutatók, általában arra a következtetésre jutottak, hogy az emberek többnyire helytelenül ítélik meg saját teljesítményüket. Például *Dunning* és *mtsai* (2003) kutatásában, amikor a diákokat arra kérték, becsüljék meg teljesítményüket egy vizsgán, a diákok legrosszabbul teljesítő negyede utólag jócskán túlbecsülte valódi teljesítményét. Arra is rávilágítottak, hogy a jobban teljesítők pontosabban meg tudták becsülni saját pontszámukat, a legjobbak pedig enyhén alulbecsülték őket.

¹ A sztenderdizációs eljárás szabályozásáról lásd *Qualifications and Curriculum Authority* (2006)

Számos elmélet tűzte ki célul a gyenge kalibráció jelenségének magyarázatát. A magabiztosság természetét kognitív pszichológusok vizsgálták, átvéve az énvédő torzítás elméletét. A kutatók azt látták, hogy a torzítást a kísérleti szituációban azért használják a résztvevők, hogy előmozdítsák, vagy fenntartsák pozitív énképüket. Ilyen megfigyelt torzítások az *átlag fölötti hatás* (Dunning, Meyerowitz és Holzberg, 2002), vagy az *optimista torzítás / indokolatlan optimizmus hatás* (pl. Armor és Taylor, 2002). A kutatások szerint az embereknek általánosan eltúlzott képzeik vannak saját képességeikről, amelyeket nem igazol objektív teljesítményük (Dunning és mtsai, 2003).

Néhány vizsgálatban a résztvevőket megkérték, hogy becsüljék meg az életükben bekövetkező pozitív és negatív események valószínűségét (Weinstein, 1980), illetve, hogy jósolják meg saját teljesítményüket egy elképzelt jövőbeli szituációban, vagy egy feladat elvégzése előtt (pl. Griffin és Tversky, 2002). Azonban ezen tanulmányokban nem vizsgálták a résztvevők valódi teljesítményét, vagy nem adtak róla visszajelzést. Az énvédő önértékeléssel foglalkozó tanulmányok elmulasztották feltárni a magabiztosság változását az aktuális teljesítményre vonatkozó visszajelzés után. Néhány tanulmányban, amelyek összevetették a résztvevők becsléseit valódi teljesítményükkel, változó eredmények születtek: míg egyes kutatók nem találtak szignifikáns összefüggést a teljesítmények és a rájuk vonatkozó becslések között (Griffin és Tversky, 2002), addig mások lényeges, szignifikáns, pozitív korrelációkat mutattak ki (pl. a kísérleti alanyok pontosan megjósolták, mennyi időre lesz szükségük egy adott feladat végrehajtásához lásd Buehler, Griffin és Ross, 1994).

Az énvédő torzítás elmélete önmagában nem képes mindezen eredményekre megfelelő magyarázattal szolgálni. Megmagyarázza, miért hajlamosak túlértékelni teljesítményüket a gyengén teljesítők, de nem mutatja meg, miért kisebb ez a különbség a jobban teljesítők esetében és miért hiányzik teljes mértékben a legjobbknál.

Az ítéletek során megnyilvánuló magabiztosság mértéke függ a társadalmi körülményektől is. Szociálpszichológusok (pl. Sherif, Sherif és Nebergall, 1965) kimutatták, hogy a laikusok hajlamosak változtatni ítéletükön nem teljesen egyértelmű ingerek esetében olyan emberek társaságában, akiket a téma szakértőinek gondolnak, vagy akik nagyon határozottan döntenek: ilyen esetekben a laikus ítélete közelít a szakértőéhez. A szakértő negatívan befolyásolja ítéletük helyességéről alkotott képüket, következőképpen magabiztosságukat, az adott helyzetben. Feltehetően egy dolgozat javításában megnyilvánuló ítéletek is számtalan bizonytalanságot tartalmaznak egy kezdő értékelő számára, az ilyen ítéleteket, valamint a kezdő értékelő hozzájuk kapcsolódó magabiztosságát tehát könnyen befolyásolhatják a szakértő értékelő megjegyzései.

A metakogníció-kutatás szintén rámutat a gyenge kalibráció okaira. A metakogníció kutatása John Flavell (1979) első ezzel kapcsolatos írása óta töretlen. A kognitív képességekkel szemben, amelyek egy probléma megoldásában játszanak szerepet, a metakogníció rámutat, hogyan jött létre a megoldás (Schraw, 1998). A szakirodalom áttekintése rávilágít, hogy a kutatók között nincs egyetértés a metakogníció, valamint az általános kogníció közötti kapcsolatot illetően. Néhányan amellet érvelnek, hogy ugyanazok a kognitív folyamatok állnak a problémamegoldás (pl. dolgozatjavítás) hátterében, mint a saját teljesítmény megítélésében az adott feladat végrehajtása során (Davidson és Sternberg, 1998). Ez magyarázatot adna rá, miért becsülik túl teljesítmé-

nyüket az alacsonyabb kognitív képességekkel bíró egyének (*Dunning és mtsai, 2003*). Mások (*Borkowski, 2000*) úgy tekintenek a metakognícióra, mint ami minőségileg teljesen elkülönül más kognitív folyamatoktól és irányítja azokat.

Schraw metakogníció-elmélete (*Schraw, 1998*) keretként szolgálhat, amely alternatív magyarázatokat nyújt a fentebb bemutatott eredményekre, valamint háttérrel is ad, amelyen az értékelők tapasztalata, az értékelés folyamata, az önértékelés mozzanata és a visszajelzés feldolgozása mind kényelmesen elhelyezhető. Valószínűleg ez a legátfogóbb elmélet, ezért hipotéziseinket és eredményeink értékelését is főként erre alapoztuk. *Schraw* (1998) nézete szerint a metakogníció két komponensből tevődik össze: *a kogníció ismeretéből*, valamint *a kogníció szabályozásából*. A kogníció ismerete három különböző metakognitív tudatosság-típust tartalmaz: deklaratív tudatosságot, azaz tudni *mit*; procedurális tudatosságot, azaz tudni *hogyan*; és kondicionális tudatosságot, azaz tudni *mikor*. A kogníció szabályozása tervezésből, monitorozásból és értékelésből áll (*Schraw, 1998*). Ezek a metakogníció azon jegyei, amelyek alapján bármilyen területen megkülönböztethető a szakértő a nem-szakértőtől.

Feltehetően a tapasztalt („szakértő”) és a tapasztalatlan („frissdiplomás”) értékelők nagymértékben különböznek egymástól metakogníciójuk tekintetében. A szakértők minden bizonnyal magasfokú deklaratív tudatossággal (tárgyi ismerettel) bírnak, minthogy megfelelő végzettséggel rendelkeznek, és általában tanítják is a tantárgyat, amit értékelnek. A kutatások azt mutatják, hogy különböző kognitív stratégiákat használnak különböző tanulói válaszok esetében (*Greatorex és Suto, 2005; Suto és Greatorex, 2008*), tehát valószínűleg magasszintű kondicionális és procedurális ismeretekkel rendelkeznek. A tapasztalatlan, frissdiplomás értékelőknek szintén rendelkezniük kell megfelelő deklaratív tudatossággal (tárgyi ismeret), azonban hiányozhat a procedurális tudásuk (pl. a hatékony értékelési stratégiák kialakítási lehetőségének hiányában), és így valószínűleg híján vannak a kondicionális metakognitív tudatosságnak is. A tapasztalatlan értékelők, kogníciós ismereteik hiányából adódó hátrányuktól függetlenül, nem feltétlenül rendelkeznek gyakorlattal a kogníció szabályozásának területén sem, egyszerűen azért, mert nem vettek még részt az értékelési folyamat tervezési, monitorozási és értékelési lépéseiben. Ebből következően a tapasztalatlan értékelőktől várható, hogy általában véve lényegesen gyengébb metakognitív képességekkel rendelkeznek, így valószínűsíthető, hogy kevésbé pontosan reflektálnak saját értékelésükre.

Bármely más kognitív képességhez hasonlóan a metakogníció is fejleszthető, egyebek között gyakorlással, ami pedig javítja a teljesítményt (jelen esetben az értékelés pontosságát) (*Koch, 2001; Dunning és mtsai, 2003*).

Az „értékelési szakértelem” kutatási projekt

A jelen tanulmányban bemutatott kutatás eredetileg egy, az értékelési szakértelmet vizsgáló nagyobb projekt része (*Suto és Nádas, 2008b*). A projekt azt vizsgálta, hogyan befolyásolja korábbi matematika és fizika GCSE dolgozatok értékelését a szakértelem más faktorokkal együtt. Fő célunk a lehetséges különbségek felderítése volt a szakértő, illetve

frissdiplomás értékelők pontossága között. Mintánkban mindkét tantárgy esetében szerepelt egy vezető értékelő, hat tanítási és értékelési tapasztalatokkal bíró („szakértő”) vizsgáztató, valamint hat frissdiplomás alapos tantárgyi ismeretekkel, de értékelésben és tanításban szerzett tapasztalatok nélkül. Minden résztvevő fizetséget kapott ugyanazon feladatsor kérdésenkénti értékeléséért, amelyet korábbi GCSE vizsgák kérdéseiből állítottunk össze. A kísérleti matematika dolgozat 20, a fizika 13 kérdést tartalmazott. Stratifikált mintagyűjtési módszerrel választottuk ki a dolgozatírók válaszait minden kérdés esetében, amelyeket lefénymásoltunk és megtisztítottunk a „valódi” értékelés során kapott pontszámoktól. Kétféle értékelési feladatsort állítottunk össze mindkét tantárgyból: egy 15 választ tartalmazó „gyakorló” és egy 50 választ tartalmazó „fő” mintát. Az értékelési eljárás minden résztvevő esetében azonosan zajlott: minden értékelő értékelési séma segítségével otthon pontozta a gyakorló tesztet, amelyről a tantárgy vezető értékelőjének irányításával visszajelzést kaptak egy egyszerű sztenderdizációs megbeszélésen. Itt kapták meg a fő mintát, amelyeket szintén otthon javítottak, de erről nem kaptak visszajelzést.

A vezető értékelők javításait vettük „helyes” vagy „igaz” válaszoknak és ezek szolgálták az adatelemzés alapjául is. Háromféle pontossági mércét alkalmaztunk: P_0 (a nyerspontszámok egyezésének mértéke a vezető értékelő és az értékelő között); az átlag valódi eltérése (MACD, azt mutatja, az értékelő átlagosan megengedőbb vagy szigorúbb vezető értékelőjénél); és abszolút átlageltérés (MABD, az értékelések eltéréseinek átlagos mértékét mutatja az értékelő és a vezető értékelő között) (a pontossági mércék részletes bemutatását lásd *Bramley, 2007*).

Meglepő módon a szakértő és frissdiplomás értékelők nagyon hasonló értékelési pontossággal dolgoztak, mind a gyakorló, mind pedig a fő tesztek esetében, mindhárom mérce szerint. A matematika gyakorló mintában a két értékelő-csoport között húsz kérdésből mindössze három esetben találtunk szignifikáns eltérést. A fő mintában egyetlen kérdésnél találtunk lényeges eltérést, amely értékelésében a kezdő értékelők némileg megengedőbbek voltak a szakértőknél, illetve a vezető értékelőnél. A fizika esetében három kérdésnél mutatkozott (tizenháromból) szignifikáns különbség a gyakorló mintában és két kérdésnél a fő mintában. Érdekes megjegyezni, hogy a szignifikáns különbségek ellenére a frissdiplomások is magas pontossági szinten értékelték minden kérdést. Mindkét csoportnál a gyakorló feladatsor hatására némi javulás történt a pontosságot illetően. Amint a további adatelemzés rámutatott, a sztenderdizációs megbeszélések, valamint az értékelési gyakorlat jótékony hatással voltak mindkét csoportra mindkét tantárgy esetében, különösen a frissen végzettekre.

Jelen tanulmány céljai és hipotézisei

Az értékelői szakértelmet feltáró kutatásunk egy további vizsgálatában, ami a jelen tanulmány tárgya, azt vizsgáltuk, hogy az értékelők hogyan észlelik saját teljesítményüket. A kalibráció és a magabiztosság vizsgálatában az értékelési folyamat három különböző pontján osztottunk ki a következő kérdések megválaszolása érdekében kérdőíveket a résztvevőknek. A kutatás céljai a következők voltak:

- 1) A kutatás első célja hogy feltárjuk, a szakértők és a kezdő értékelők mennyire magabiztosan ítélik meg értékelési pontosságukat a sztenderdizációs megbeszélést megelőzően. Egyrészt a metakognitív elméletre támaszkodva, másrészt tudván, hogy a kezdőket általánosan kevésbé akkurátusnak tartják, mint a szakértőket, két hipotézist véltünk tarthatónak: a) a frissdiplomások tudatában vannak metakognitív képességeik hiányosságainak, ezért magabiztosságuk alacsonyabb szintet fog mutatni a szakértőkéhez viszonyítva a gyakorló értékelés után; valamint b) a kezdők nincsenek tudatában hátrányuknak, így a gyakorló tesztek után hasonló mértékű magabiztossággal találkozhatunk valamennyi értékelőnél. Az első hipotézis tűnik valószínűbbnek, mint hogy a kezdő értékelők tudták, hogy szakértők is részt vesznek a vizsgálatban.
- 2) A szakértők és frissdiplomások magabiztosságában az értékelési folyamat során végbemenő változások feltárása. A metakognitív elmélet azt vetíti előre, hogy a szakértők magabiztossága az értékelés folyamatában mindvégig magas szinten lesz, illetve némi emelkedést is mutathat, minthogy az értékelési gyakorlat, valamint az egyes vizsgakérdésekre vonatkozó visszajelzések fejleszthetik metakognitív képességeiket is. A folyamat során elsajátított értékelői gyakorlatnak köszönhetően feltételezhető, hogy a kezdők magabiztossága emelkedő tendenciát fog mutatni. Ezért az egymást követő kérdőíveken a frissdiplomások valószínűleg egyre magasabb szintű magabiztosságot fognak mutatni, addig a szintig, amikor magabiztosságuk eléri a szakértők szintjét. Esetleges alternatívaként a metakognitív elmélete szerint a frissdiplomások magabiztossága csökkenhet is (a sztenderdizációs megbeszélést követően a második kérdőíven) két okból kifolyólag: a kezdő értékelők ítéleteit befolyásolhatja a szakértő értékelők jelenléte. Jóllehet tudtak a vizsgálatban való részvételükről, a megbeszélés során azonban olyan új értelmezési keretet adhattak a frissdiplomásoknak. Másodszor a vezető értékelőtől ezen a ponton kapnak visszajelzést, ami hatására megeshet, hogy át kell gondolniuk a gyakorlatban nyújtott pontosságukat. Ez azt is sejteti, hogy a kezdők és a szakértők magabiztossága egyöntetűen a fő mintára vonatkozóan lesz a legmagasabb, valamint, hogy nagyon hasonló lesz a két csoport tekintetében.
- 3) A kezdők és a szakértők kezdeti kalibrációs mintázatának feltárása, valamint a két csoport közötti szignifikáns különbségek feltárása. A metakognitív elmélet azt jósolja, hogy a kezdők gyenge kalibrációs készségekkel rendelkeznek a procedurális és kondicionális metakognitív tudatosság hiányából adódóan, míg a szakértők előző értékelési tapasztalataik és pontosságukra kapott visszajelzések birtokában jobb teljesítményt nyújtanak.
- 4) A résztvevők kalibrációs képességeinek fejlődésének feltárása az értékelési folyamat során. A metakognitív elmélet szerint minden résztvevő, de különösen a frissdiplomások saját teljesítményükkel kapcsolatos rálátásának emelkedő tendenciát kellene mutatnia az egymást követő kérdőíveken, mert a folyamat végére már gyakorlat és visszajelzés is lesz a birtokukban, valamint metakognitív képességeiket is fejleszthetik a kérdőívek kitöltéséhez szükséges áttekintés során saját teljesítménybecslésükkel kapcsolatban.

Amint már korábban említettük, a szakirodalom alapján azt mondhatjuk, néhány kutató úgy látja, a metakognitív képességek ugyanazokat a kognitív folyamatokat hozzák működésbe, amelyeket a problémamegoldás maga, mások szerint viszont a metakognitív

felsőbb, rendező elve más kognitív folyamatoknak. Mivel értékelési szakértelemmel foglalkozó projektünk első vizsgálata szerint a frissdiplomások és a szakértők nagyon hasonló teljesítményt nyújtottak az értékelési pontosság terén (Suto és Nádas, 2008b), feltételezhetjük, hogy nem az alapvető kognitív képességeik okolhatók a metakognitív képességeikben megjelenő különbségekért (ha egyáltalán adódnak ilyen különbségek). Amennyiben helyes az érvelésünk, bármely különbség a két értékelő-csoport metakognícióját illetően a fentebb megnevezett folyamatok különbségeiből származtatható (a *procedurális tudatosság*, tudni mit, és a *kogníció szabályozása*, azaz tervezés, monitorozás és értékelés különbségeiből) inkább, mintsem a kognitív képességek eltéréseiből. Ez azt támasztaná alá, hogy a metakogníció nem azonos más kognitív folyamatokkal.

Módszer

Minta

A minta 26 értékelőből állt. Tantárgyanként hat szakértő értékelő (tárgyi ismeretekkel, a kiválasztott vizsgadolgozat értékelésében szerzett tapasztalattal, valamint tanítási gyakorlattal), hat frissdiplomás értékelő (tantárgyi ismeretekkel, de tanítási és értékelési tapasztalatok nélkül) és egy igen tapasztalt vezető értékelő vett részt vizsgálatunkban.

Eljárás

A kutatás megkezdésekor minden résztvevő kapott egy értesítő levelet, amely leírta, hogy mind szakértő, mind pedig frissdiplomás értékelők részt vesznek a kutatásban, valamint minden értékelő ugyanazokat a „gyakorló” és „fő” teszteket fogja értékelni. Az értékelők három alkalommal töltötték ki kérdőíveket: (1) a sztenderdizációs megbeszélés elején, a gyakorló feladatsor (15 kérdés) értékelését követően; (2) a megbeszélést követően; és végül (3) a fő teszt (50 kérdés) értékelését követően.

Az első (a megbeszélés kezdetekor kitöltött) és második (a megbeszélés végeztével kitöltött) kérdőívben a következő kérdést tettük fel: „Megérzése szerint, mennyire pontosan értékelte az első – gyakorló – kérdéssor tanulói válaszait?”

A harmadik kérdőívben (a fő feladatsor értékelése után) minden értékelőt megkérdeztünk arról, hogy „Megérzése szerint, mennyire pontosan értékelte a második – fő – kérdéssor tanulói válaszait?”

A válaszokat a következő 5 fokozatú skálán kellett megadni: (1) Nagyon pontatlanul, (2) Pontatlanul, (3) Fogalmam sincs, (4) Pontosan, (5) Nagyon pontosan.

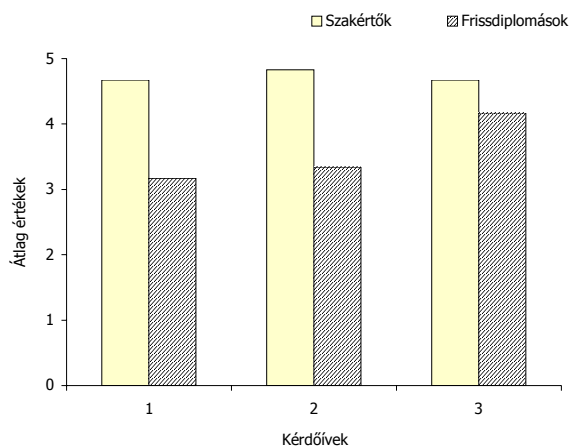
Eredmények

Az adatok eloszlásának ellenőrzése után kiszámoltuk a magabiztossági átlagokat és t-próbát, illetve Mann-Whitney-féle u-próbát alkalmaztunk a két értékelő-csoport esetleges különbségeinek elemzésére. Megvizsgáltuk a Pearson- és Spearman-féle korrelációk

koefficienseket annak felderítésére, van-e kapcsolat az értékelési pontosság és a magabiztosság adatai között.

A kezdő és a szakértő értékelők magabiztosságának elemzése

Az 1. ábra a kezdő és szakértő matematika értékelők magabiztosság-értékelésének átlagait mutatja mindhárom kérdőívre vonatkoztatva. A t-próbák tanulsága szerint a két csoport magabiztosság-értékelése a gyakorló tesztre vonatkozóan szignifikánsan különbözik az első ($t=4,02$, $p<0,01$) és második ($t=2,87$, $p<0,05$) kérdőív esetében, amelyeknél a kezdők szignifikánsan alacsonyabbra értékelték saját pontosságukat. Ez a különbség a harmadik kérdőívben teljesen eltűnt ($t=1,86$, $p>0,05$); a két értékelő-csoport nagyon hasonlóan ítélte meg a fő feladatsor értékelésének pontosságát. Változást a magabiztosságban csak a frissdiplomás csoport esetében találtunk, akik magabiztossága szignifikánsan növekedett az első és a harmadik kérdőív kitöltése között ($t=-3,83$, $p<0,05$).



1. ábra

Szakértő és frissdiplomás matematika értékelők magabiztosság-értékelésének átlagait bemutató grafikon

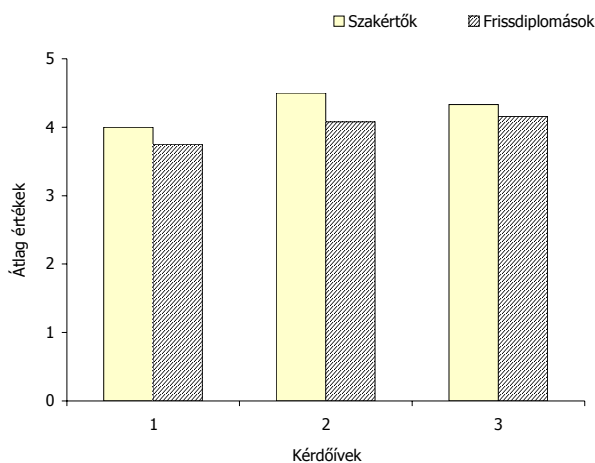
Kérdőív 1: „Megérzése szerint, mennyire pontosan értékelt az első – gyakorló – kérdéssor tanulói válaszait?”;

Kérdőív 2: „Megérzése szerint, mennyire pontosan értékelt az első – gyakorló – kérdéssor tanulói válaszait?”;

Kérdőív 3: „Megérzése szerint, mennyire pontosan értékelt a második – fő – kérdéssor tanulói válaszait?”

A 2. ábra a fizika értékelők magabiztosság-értékelésének átlagait mutatja, összehasonlítva a szakértők és kezdők átlagait. A matematika értékelőivel ellentétben nem mutatkozott szignifikáns különbség a két értékelő-csoport között egyik kérdőív esetében sem.

A magabiztosság és a teljesítménybecslés pontosságának kutatása az angol GCSE vizsgák értékelőinél



2. ábra

Szakértő és frissdiplomás fizika értékelők magabiztosság-értékelésének átlagait bemutató grafikon

Kérdőív 1: „Megérzése szerint, mennyire pontosan értékelt az első – gyakorló – kérdéssor tanulói válaszait?”;

Kérdőív 2: „Megérzése szerint, mennyire pontosan értékelt az első – gyakorló – kérdéssor tanulói válaszait?”;

Kérdőív 3: „Megérzése szerint, mennyire pontosan értékelt a második – fő – kérdéssor tanulói válaszait?”

A szakértők és a frissdiplomások kalibrációs pontosságának elemzése

Annak megállapítására, hogy az értékelőknek milyen mértékű rálátásuk van saját értékelési teljesítményükkel kapcsolatban, megkíséreltük összevetni a két értékelő-csoport magabiztosság-adatait három értékelési pontosságot mérő eszközünk (P_0 , MAbD, MAbD) által mért adatokkal mind a gyakorló, mind pedig a fő minta esetében.

A matematika esetében a Pearson-féle korrelációs koefficiens tanúsága szerint sem a szakértő, sem pedig a kezdő értékelők nem tudták helyesen megbecsülni saját értékelésük pontosságát egyik tesztnél sem. A magabiztosság nem mutatott szignifikáns kapcsolatot egyik mérőeszköz eredményével sem. A koefficiens értékek a következőképpen alakultak: a szakértőknél: $r=-0,46$, $p=0,36$ az első kérdőív esetén; $r=-0,29$, $p=0,58$ a második kérdőív esetén; és $r=-0,47$, $p=0,34$ a harmadik kérdőív esetén; a frissdiplomásoknál: $r=0,43$, $p=0,40$ az első kérdőív esetén; $r=0,02$, $p=0,97$ a második kérdőív esetén; és $r=0,46$, $p=0,35$ a harmadik kérdőív esetén.

A fizika tantárgy esetében a Spearman- és Pearson-féle korrelációs koefficiensek jeleztek némi szignifikáns korrelációt. A szakértők fő értékelését követő magabiztossága (harmadik kérdőív) szignifikánsan pozitívan korrelált a fő minta P_0 átlagával ($r=0,83$, $p<0,05$) és erősen negatívan a MAbD átlagával ($r=-0,86$, $p<0,05$). Ellenben a frissdiplomások magabiztossága szignifikáns negatív összefüggést mutatott a P_0 átlagértékekkel ($r=-0,81$, $p<0,05$) viszont a sztenderdizációs megbeszélést követően (második kérdőív) pozitívan korrelált a MAbD átlagokkal ($r=0,86$, $p<0,05$). Mindkét korreláció azt mutatja,

hogy a szakértők minél pontosabban értékelték a mintát, annál magasabb szintű magabiztosságról adtak számot. A fő feladatsor esetén magas szintű rálátást mutattak értékelési pontosságukkal kapcsolatban. A kezdőknél a gyakorló feladatsor esetében éppen fordított a helyzet: minél magabiztosabbak voltak, annál kevésbé bizonyult pontosnak értékelésük (a két mérőeszköz szerint). Az 1. táblázat foglalja össze ezeket az eredményeket.

1. táblázat. A magabiztosság és az értékelési pontosság kapcsolatának összefoglaló táblázata

| | <i>Az 1. kérdőíven jelölt magabiztosság szignifikánsan korrelál-e a gyakorló teszt értékelésének pontosságával?</i> | <i>A 2. kérdőíven jelölt magabiztosság szignifikánsan korrelál-e a gyakorló teszt értékelésének pontosságával?</i> | <i>A 3. kérdőíven jelölt magabiztosság szignifikánsan korrelál-e a fő feladatsor értékelésének pontosságával?</i> |
|----------------------------|---|--|---|
| Matematika szakértők | nem | nem | nem |
| Matematika frissdiplomások | nem | nem | nem |
| Fizika szakértők | nem | nem | pozitív korreláció |
| Fizika frissdiplomások | nem | negatív korreláció | nem |

Diszkusszió

Mindent egybevetve, vegyes eredményre jutottunk: hipotéziseinket csak részben igazolták adataink, és nagyon különböző mintázatot találtunk a matematika- és fizika-értékelők magabiztosságában, illetve teljesítménybecslésük pontosságában.

Első célunk az volt, hogy feltárjuk a szakértő és kezdő értékelők magabiztosságát a sztenderdizációs megbeszélés előtt. Minden szakértő értékelő magas szintű kezdeti magabiztosságról adott számot. A matematika szakértők átlaga némileg magasabb volt a fizika-értékelők mindkét csoportjánál. Úgy tűnik, mindkét hipotézisünk – nevezetesen, hogy a frissdiplomások hasonló szintű magabiztossággal rendelkeznek, mint a szakértők, illetve, hogy a gyakorló feladatsor esetén kevésbé lesznek magabiztosak – egyenként teljesült a két különböző tantárgyat értékelő csoport esetén: a matematikát értékelő frissdiplomások szignifikánsan kevésbé voltak magabiztosak, mint e terület szakértő értékelői, amely eredmény tükrözi metakognitív és értékelési képességeik hiányára vonatkozó elvárásainkat. A fizikát értékelő frissdiplomások azonban nem mutattak eltérést a szakértőktől magabiztosságuk tekintetében, ami a metakognitív keretrendszerben azt jelentheti, hogy nem számoltak saját tapasztalatlanságukkal. Amennyiben viszont figyelembe vesszük kezdő fizika értékelők magas szintű értékelési pontosságát, azt mondhatjuk, magabiztosságuk szintje összhangban van aktuális teljesítményükkel. Végül, a matematika és a fizika csoport gyakorló teszt után tanúsított magabiztossága közötti eltérés okaira nem derült fény.

Második számú célunk annak feltárása volt, hogyan változik az értékelők magabiztossága az értékelési folyamat alatt. A metakognitív elmélet magyarázattal szolgálhat arra a megfigyelésre, hogy a szakértők magabiztossága következetesen magas volt; azonban magabiztosságuk szintjében nem észleltünk emelkedést az értékelés folyamata alatt. Jóllehet a metakognitív elmélet alapján enyhe növekedést vártunk volna, a vizsgálatban foglalt értékelés mennyisége talán nem bizonyult elégnek a metakognitív képességek továbbfejlesztéséhez, esetleg a résztvevők ezen képességei már a kutatás megkezdésekor maximum szinten voltak.

Feltevésünknek megfelelően a matematika szakos frissdiplomások növekvő szintű magabiztosságról adtak számot, egészen addig, míg az első és második kérdőívben diagnosztizált szignifikáns különbség a szakértők és a frissdiplomások között teljesen eltűnt a fő feladatsor értékelését követően. Ezzel ellentétben a fizika szakos kezdő értékelők ugyanolyan magabiztosak voltak az értékelés folyamatának egészében, mint tapasztaltabb kollégáik. Ez igen meglepő, minthogy a frissdiplomásoknak saját teljesítményük értékelése során figyelembe kellene venniük előzetes tapasztalataik hiányát (amelyet már a gyakorló teszt értékelésekor sem tettek). Ugyanakkor szinte ugyanolyan pontosan értékelték, mint a szakértők, tehát magabiztosságuk helyénvalónak bizonyult, jóllehet váratlan, magas értékelési pontosságukhoz hasonlóan.

Az adatok nem támasztották alá azon feltevésünket, mely szerint a frissdiplomások magabiztossága csökkenni fog a sztenderdizációs megbeszélést követően. Úgy tűnik, sem az új szociális referencia (amelyet a szakértők jelenlététől vártunk), sem a visszajelzések nem befolyásolták a frissdiplomások magabiztosságát egyik tantárgy esetében sem. Valamennyi frissdiplomás magabiztossága elérte a maximum szintet a fő teszt értékelését követően, amikor minden különbség (ha volt egyáltalán) eltűnt közöttük és a szakértők között.

Harmadik célunk a résztvevők saját teljesítményét illető kezdeti rálátásának, kalibrációjának feltárása volt, amelyet a magabiztosság és a pontosság közötti potenciális összefüggések mutattak meg. Meglepő módon a megbeszéléseken kapott visszajelzéseket megelőzően egyik értékelő kalibrációja sem bizonyult jónak. Ez kiváltképp érdekes a szakértő értékelők esetében, mivel a metakognitív elmélet éppen az ellenkezőjét jósolta, figyelembe véve előzetes tapasztalataikat saját értékelési pontosságuk megítélésében. Úgy tűnik tehát, hogy a különböző vizsgakérdések értékelésében szerzett előzetes tapasztalatok, valamint a saját értékelésre vonatkozó reflexiók nem feltétlenül generalizálódnak, amikor új itemek értékeléséről vagy új értékelés pontosságának meghatározásáról van szó.

Végül pedig a kalibráció változásait kíséreltük meg feltárni a négy értékelő-csoport esetében az értékelés folyamata során. A metakognitív elmélet szerint minden csoport kalibrációjának, de különösen a frissdiplomás csoportokénak, fejlődő tendenciát kellett volna mutatnia az egymást követő kérdőíveken. A matematika esetében, meglepő módon, egyik csoport metakognitív teljesítménye sem javult a tapasztalatok szerzésével, ahogyan egyik csoport becslése sem volt pontos saját teljesítményére vonatkozóan sem a gyakorló feladatsornál a megbeszélést követően, sem a fő minta esetén. A matematika-értékelők adatai tehát nem igazolták a metakognitív hipotézist.

Várakozásaink a fizika esetén is csak részben teljesültek: a tapasztalt értékelők mutattak némi rálátást teljesítményükkel kapcsolatban, azonban csak a fő minta esetében.

Így ez esetben úgy tűnik, a feltevést, mely szerint a metakogníció gyakorlással fejleszhető, adataink megerősítették. Meglepő módon, a fizika szakos frissdiplomások becslése és valódi teljesítménye között szignifikáns negatív korrelációt találtunk a gyakorló feladatsor esetén. Ez pedig látszólag az énvédő torzítás elméletét támasztja alá, amely előre jelezte ezt a túlzott optimizmust. Az elmélet azonban minden csoport számára ezt jósolta, amelyet adataink nem erősítettek meg.

A fő minta értékelésekor mind a tapasztalt, mind pedig a kezdő értékelők igen nagy pontosságot értek el, ezért metakognitív képességeik különbségét nem tulajdoníthatjuk kognitív képességeik különbségének. Valójában igen valószínű, hogy a kogníció szabályozása, illetve a procedurális tudás hiányosságai felelősek az említett eltérésért. Ez megvilágítja a kogníció és metakogníció kapcsolatának természetét is: a tény, hogy a frissdiplomás fizika értékelők a szakértőkhöz hasonlóan teljesítettek egy kognitív képességeket igénybevevő feladatban, ugyanakkor a két csoport metakogníciója különböző mintázatot mutatott, azt sugallja, hogy a két folyamat lényegét tekintve nem ugyanaz a jelenség. Természetesen ennek részletes kifejtéséhez további empirikus vizsgálatokra van szükség.

Korlátok

Mint minden kutatásnak, a mi vizsgálatunknak is voltak korlátai. Az egyik legnyilvánvalóbb, hogy kutatásunk kis létszámú csoportokkal dolgozott, amelyek tagjai több változót tekintve is eltértek egymástól: a szakértő értékelők mind tanítási, mind pedig értékelési tapasztalatokkal rendelkeztek, míg a frissdiplomások mind fiatal értelmiségiek voltak. Emellett, számos frissdiplomás a Cambridge-i Egyetemen végzett, amely már önmagában hatással lehet az eredményekre; az itt végzettek például akadémiusabb szemlélettel rendelkezhetnek, esetleg lelkiismeretességben, illetve határozottságban térhetnek el a más intézményi háttérrel rendelkezőknél. A további kutatásban a szakértelmet és az értékelői háttérrel tekintve szélesebb körű mintára lesz szükség.

További korlátot jelentett, hogy a vizsgálat mindössze kettő, természetükben hasonló vizsga-feladatsort használt. Más tantárgyak bevonása más eredményre vezethetett volna. Szintén fenntartásokra ad okot, hogy nem tudjuk, az értékelők mennyire vették komolyan kérdőíveinket, azaz, hogy fordítottak -e időt arra hogy saját értékelési teljesítményüket megfontolva töltsék ki, vagy különösebb önreflexió nélkül adtak meg adatokat. Ez irányú bizonytalanságunkat növeli, hogy „kísérleti” vizsgákat használtunk, csak és kizárólag kutatási céllal, tehát az adott pontszámok semmilyen módon nem voltak hatással senki életére. „Éles” értékelési helyzetben esetleg más értékeket kaptunk volna a résztvevők magabiztosságában és rálátásában. Végezetül pedig másféle korlátot jelentett, hogy a résztvevők egyszerre szereztek gyakorlatot az értékelésben és hajtottak végre metakognitív feladatokat, így a vizsgálat nem tette lehetővé hatásaik elkülönített elemzését. A további kutatásban ezeket el kell választani egymástól.

Konklúzió és további kutatás

Különböző tantárgyak értékelői nagyon eltérő mértékű magabiztosságról és teljesítménybecslési pontosságról adnak számot. A frissdiplomás matematika értékelők szignifikánsan alacsonyabb magabiztosság értékeket adtak meg a gyakorló teszt esetén, mint a szakértők, azonban a fő mintán már nem volt különbség a két csoport között. A kezdő fizika értékelők viszont ugyanolyan magabiztosak voltak a folyamat egésze alatt, mint szakértő kollégáik. Általában véve, az értékelők magabiztossága állandó szinten volt a folyamat során, egyedül a frissdiplomás matematika értékelők magabiztossága mutatott fejlődést a gyakorló teszt első értékelése és a fő teszt értékelése között.

Néhány értékelő tanúsított némi rálátást értékelési teljesítményére vonatkozóan, azonban ez nem volt következetes és még a szakértők kalibrációja sem volt mindig megfelelő. A matematika értékelők nem mutattak semmilyen rálátást értékelési pontosságukkal kapcsolatban sem a gyakorló, sem a fő teszt esetében. A tapasztalt fizika értékelők korrekt teljesítménybecslést mutattak a fő feladat esetében, a frissdiplomások pedig szignifikáns negatív korrelációt produkáltak becslött és valós teljesítményük között, a gyakorló feladatsornál.

Vegyes eredményeink miatt egy elmélet sem képes teljesen megmagyarázni adatainkat. Azonban úgy tűnik, legtöbbször, ha nem is minden eredményünk értelmezhető a metakognitív elmélet keretrendszerében, azaz jelen tanulmány egyúttal a kogníció és metakogníció kapcsolatának természetére irányuló empirikus vizsgálatként is szolgál. A tapasztalt és kezdő fizika értékelők kalibrációjában mutatkozó különbségek nem tükrözték az értékelési pontosságukban mutatkozó általános hasonlóságot, ezért metakognitív képességeik eltérése minden bizonnyal procedurális és kondicionális tudatosságuk különbségéből ered, nem pedig kognitív képességeikből. Ez azt sugallja, hogy a kogníció és a metakogníció minőségileg eltérő folyamatokban nyilvánul meg. Tisztázatlan maradt továbbá, miért mutatkozott ekkora különbség a matematika és fizika értékelők magabiztosságában és teljesítménybecslésük pontosságában.

Amint azt a bevezetésben említettük, kutatásunknak gyakorlati haszna lehet a sztenderdizációs megbeszéléseken, ahol a vezető értékelők megtárgyalják csoportjaikkal azokat a kérdéseket, amelyek bizonytalanságot okoztak az értékelőknek. Jelen vizsgálatunk azonban rámutatott, hogy az értékelőknek, különösen a matematika esetében, nincs rálátásuk saját pontosságukra, így nem tudják megmondani, mely kérdéseket kellene megbeszélni a találkozókon. Ez kiküszöbölhető lenne számítógépes értékeléssel, amely esetben a sztenderdizációs eljárásban az értékelők azonnali visszajelzést kapnának pontosságukról, azáltal is javítva teljesítménybecslésük pontosságát; vagy olyan kvalitatív kutatásokkal (például a Kelly-féle szereprepertoár-teszt alkalmazásával) amelyben a vezető értékelők is részt vennének további információkat szolgáltatva arról, mely tulajdonságok tesznek egy vizsgakérdést nehezebben értékelhetővé másoknál (lásd *Suto és Nádas*, 2008a).

Az értékelők metakogníciójának kutatását kiterjesztettük egy folyamatban lévő utánkövetéses vizsgálatra is, amelyben az első tanulmány számos korlátját sikerült kiküszöbölni egy kifinomultabb kutatási terv segítségével. Ebben a kísérleti értékelő vizsgálatban azt figyeljük, miként értékel több, mint nyolcvan különböző tapasztalatokkal rendel-

kező értékelő üzleti tanulmányok GCSE, illetve biológia nemzetközi GCSE (IGCSE) dolgozatokat. Az értékelők metakognícióját, valamint személyiségjegyeit kibővített kérdőívekkel vizsgáljuk. Az említett vizsgálat adatainak elemzése folyamatban van, eredményeit 2008-ban tervezzük publikálni.

Köszönetnyilvánítás

Jelen kutatás az Oxford, Cambridge és RSA Examinations (OCR) által szervezett vizsgákon alapul és a Cambridge Assessment anyagi támogatásával jött létre.

Irodalom

- Annor, D. A. és Taylor, S. E. (2002): When Predictions Fail: The Dilemma of Unrealistic Optimism. In: Gilovich, T., Griffin, D. és Kahneman, D. (szerk.): *Heuristics and Biases: The Psychology of Intuitive Judgement*. Cambridge University Press, Cambridge. 334–347.
- Borkowski, J. G. (2000): *The assessment of executive functioning*. Paper delivered at the annual convention of the American Educational Research Association, New Orleans.
- Bramley, T. (2007): Quantifying marker agreement: terminology, statistics and issues. *Research Matters: A Cambridge Assessment Publication*, 4. sz. 22–28.
- Buehler, R., Griffin, D. és Ross, M. (1994): Exploring the “planning fallacy”: Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67. sz. 366–381.
- Davidson, J. E. és Sternberg, R. J. (1998): How metacognition helps. In: Hacker, D. J., Dunlosky, J. és Graesser, A. C. (szerk.): *Metacognition in Educational Theory and Practice*. London: Lawrence Erlbaum Associates, Publishers, Hillsdale. 47–68.
- Dunning, D., Heath, C. és Suls, J. M. (2004): Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5. 3. sz. 69–106.
- Dunning, D., Johnson, K., Ehrlinger, J. és Kruger, J. (2003): Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 83–87.
- Dunning, D., Meyerowitz, J. A. és Holzberg, A. D. (2002): Ambiguity and self-evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments and Ability. In: Gilovich, T., Griffin, D. és Kahneman, D. (szerk.): *Heuristics and Biases: The Psychology of Intuitive Judgement*. Cambridge University Press, Cambridge. 324–333.
- Flavell, J. H. (1979): Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34. sz. 906–911.
- Greatorex J., Nádas, R., Suto, W. M. I. és Bell, J. F. (2007): Exploring how the cognitive strategies used to mark examination questions relate to the efficacy of examiner training. Paper presented at ECER Ghent - Belgium, September.
- Greatorex, J. és Suto, W. M. L. (2005): *What goes through a marker's mind? Gaining theoretical insight into the A-level and GCSE marking process*. A report of a discussion group at the Association for Educational Assessment – Europe, Dublin.
- Gliffin, D. és Tversky, A. (2002): The Weighing of Evidence and the Determinants of Confidence. In: Gilovich, T., Griffin, D. és Kahneman, D. (szerk.) *Heuristics and Biases: The Psychology of Intuitive Judgement*. Cambridge University Press, Cambridge. 230–249.
- Koch, A. (2001): Training in metacognition and comprehension of physics texts. *Science education*, 85. 6. sz. 758–768.

A magabiztosság és a teljesítménybecslés pontosságának kutatása az angol GCSE vizsgák értékelőinél

- Maki, R. H. (1998): Test predictions over text material. In: Hacker, D. J., Dunlosky, J. és Graesser, A. C. (szerk.): *Metacognition in Educational Theory and Practice*. London: Lawrence Erlbaum Associates Publishers, Hillsdale, 117–144.
- Qualifications and Curriculum Authority (2006): *GCSE, GCE, VCE, GNVQ and AEA Code of Practice 2006/7*. London: Qualifications and Curriculum Authority.
- Schraw, G. (1998): Promoting General Metacognitive Awareness. *Instructional Science*, 26. sz. 113–25.
- Sherif, C., Sherif, M. és Nebergall, R. (1965): *Attitude and attitude change: The social judgement-involvement approach*. Saunders, Philadelphia.
- Stankov, L., és Kleitman, S. (2008). Processes on the Borderline Between Cognitive Abilities and Personality: Confidence and Its Realism. In Boyle, G.J., Matthews, G. és Saklofske, D. (szerk.) (pp. 541-555). *The Handbook of Personality Theory and Testing*. Sage Publications.
- Suto, W. M. I. és Greatorex, J. (2008): A quantitative analysis of cognitive strategy usage in the marking of two GCSE examinations. *Assessment in Education: Principles, Policy and Practice* 15. 1. sz. 73-90.
- Suto, W. M. I. és Nádás, R. (2007): The 'Marking Expertise' projects: Empirical investigations of some popular assumptions. *Research Matters: A Cambridge Assessment Publication*, 4. sz. 2–5.
- Suto, W. M. I. és Nádás, R. (2008a): Why are some GCSE examination questions harder to mark accurately than others? Using Kelly's Repertory Grid technique to identify relevant question features. *Research Papers in Education*
- Suto, W. M. I. és Nádás, R. (2008b): What determines GCSE marking accuracy? An exploration of expertise among maths and physics markers. *Research Papers in Education*
- Vidal Rodeiro, C. L. (2007): *Uptake of GCSE subjects 2000-2006*. Statistics Report Series No. 4., http://www.cambridgeassessment.org.uk/ca/Our_Services/Research/Statistical_Reports
- Weinstein, N. D. (1980): Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39. sz. 806–820.

Nádas Rita és Irenka Suto

ABSTRACT

RITA NÁDAS AND IRENKA SUTO: AN EXPLORATION OF SELF-CONFIDENCE AND INSIGHT INTO MARKING ACCURACY AMONG GCSE MATHS AND PHYSICS MARKERS

GCSE examinations are high-stakes national assessments taken by 16-year-olds in England. They are administered and assessed by national exam boards like Cambridge Assessment, and emphasis is placed on ensuring marking accuracy and reliability. Cambridge Assessment is conducting a series of studies exploring markers' cognitive thought processes in order to improve the quality of its assessments. Our research investigates markers of differing expertise, and includes an exploration of markers' insights into the quality of their own marking. Research into self-assessment has revealed that people often lack insight into their competence in performing cognitively demanding tasks, and suggests that practice and feedback enhance insight. This empirical study aimed: (1) to investigate how markers' self-confidence changes during the course of the GCSE marking process; and (2) to explore how much insight markers have into their quality of marking; (3) to identify any differences in insight between experienced and inexperienced markers. Candidates' responses from past UK GCSE maths and physics examination questions were marked by groups of experienced and inexperienced (graduate) markers. Questionnaires asked markers about their perceived accuracy on three occasions during the marking process. We found that marking accuracy varied among individual markers, and that groups showed differing patterns of insight, with some significant correlations between performance and insight. Generally, however, it appears likely that markers' own perspectives are poor indicators of their accuracy.

Magyar Pedagógia, **107**. Number 3. 169–184. (2007)

Levelezési cím / Address for correspondence: Nádas Rita, Research Officer, Research Division, Cambridge Assessment, 1 Hills Road Cambridge CB1 2EU United Kingdom,
E-mail: nadas.r@cambridgeassessment.org.uk