

Markmyprofessor: Minél több, annál jobb?

Markmyprofessor: the More the Better?

M. KISS, V. B. SÁTOR, E. KONTOR, A. I. KUN

Debreceni Egyetem, Gazdaságtudományi Kar, Marketing Tanszék, kiss.marietta@econ.unideb.hu

Független kutató, sator.viktor@gmail.com

Debreceni Egyetem, Gazdaságtudományi Kar, Marketing Tanszék, kontor.eniko@econ.unideb.hu

Debreceni Egyetem, Gazdaságtudományi Kar, Emberi Erőforrás Menedzsment Tanszék,

kun.andras.istvan@econ.unideb.hu

Absztrakt. Az oktatók hallgatók általi értékelése a felsőoktatásban nemzetközileg és Magyarországon is aktuális téma, részben a felsőoktatás tömegesedése, részben piacosodása miatt. Az értékelési módszerek egyike az internetes alapú önkéntes véleményezés, melyet az Amerikai Egyesült Államokban a ratemyprofessors.com indított el, és amelynek hazai megfelelője a markmyprofessor.com. Az amerikai oldallal kapcsolatos elemzések a számos kritika egyikeként megállapították azt is, hogy ezen értékelési mód megbízhatósága jelentős mértékben függ az egyes oktatókra leadott értékelések számától. Tanulmányunkban ez utóbbi következtetést vizsgáljuk meg az említett hazai szolgáltatást alapul véve. Fő eredményünk egyrészt az, hogy a ratemyprofessors.com-ra tett előbbi megállapítás hazánkban is megállja a helyét, másrészt pedig az, hogy a magasabb válaszadási gyakoriság egyben alacsonyabb átlagos értékeléssel is együtt jár.

Abstract. Student evaluation of faculty members is a current topic both internationally and in Hungary, partly due to the massification and partly due to the marketization of higher education. One of the evaluation methods is the internet-based voluntary rating which was started by ratemyprofessors.com in the U.S.A. and whose Hungarian counterpart is markmyprofessor.com. Research studies on the U.S. website – among many other critics – established that reliability of this evaluation method highly depends on the frequency of ratings: if an individual tutor is evaluated by more students, then the standard deviation of the ratings is smaller. The previous conclusion about the U.S. website is tested in our study based on Hungarian data. Our main result is that many of the experiences about markmyprofessor.com echo the previous findings about ratemyprofessors.com. Beside this in the Hungarian sample the higher response rate is associated with lower average ratings.

Bevezetés

A felsőoktatási intézmények – a felsőoktatás tömegesedése és piacosodása miatt is kiéleződő verseny hatására – egyre nagyobb hangsúlyt fektetnek a hallgatók elégedettségére. Utóbbi mérésére számtalan lehetőség áll az intézmények rendelkezésére, a formális, éves-féléves hallgatói értékelésektől (akár offline, akár online formában) kezdve a strukturált mélyinterjúkon át az informális értékelésekig (beszélgetések, oktatókat értékelő internetes oldalak). Vizsgálatunkban az informális online értékelő

oldalakat vettük górcső alá, melyek itthon és külföldön is egyre nagyobb népszerűségnek örvendenek [1], és számos kutatás foglalkozik létjogosultságukkal, esetleges problémáikkal.

1. Elméleti háttér

A fő vita, amely az oktatók hallgatói értékelését övezi, az értékelés eredményeinek felhasználási céljaiból fakad. Külföldi egyetemeken az éves formális hallgatói értékelések mellett akár az online oldalakon megjelenő információ is befolyásolhatja az oktatók karrierjét [2], [3], [4]. Magyarországon az online értékelések hivatalosan nem befolyásolják az oktatók pályáját, azonban mivel az informális értékelő felületeket bárki, bárhonnán elérheti és referenciapontként használhatja (akár valós az ezáltal kialakult kép, akár nem), mindenképpen érdemes foglalkozni a témával.

Az oktatók hallgatók általi értékelésének alapvető célja Otto et al. [3] szerint nem az oktató teljesítményének mérése (majd ennek az információnak a felhasználása az elbocsátásról, előléptetésről, kinevezésről vagy fizetésemelésről szóló döntések során), hanem az, hogy a hallgatók tanulási hatékonyságát fokozzák (például azzal, hogy segítik a nekik megfelelő oktató megtalálását a szabad oktatóválasztás során). Szerintük ugyanis az oktatók számára annak kellene elsődlegesnek lenni, hogy a diákok számára megkönnyítsék a tanulást, ezzel is növelve az intézményük által nyújtott szolgáltatás minőségét, nem pedig elégedettségük közvetlen növelése. A hallgatók és az oktatók értelmezése ebből kifolyólag eltérhet arról, hogy milyen is a jó kurzus (és ebből következően hogy milyen a jó oktató). A hallgatók rövid távú elégedettségét ugyanis jobban befolyásolhatja az, hogy mennyi időt kell az adott kurzus abszolválására fordítani, mint az elérhető tudás szintje, ami csak később hasznosul ([5] alapján, magyarázat a szerzőktől).

Az online értékelő oldalokról származó értékelések Cashin [2], [6], Greenwald és Gillmore [7], Wilson [8], valamint Liaw és Goh [9] szerint az oktatók személyiségéből, karizmájából és engedékenységből fakadóan elfogultak, ezért nem használhatók sem az oktatói teljesítmény, sem a hallgatók tanulásának mércéjeként. Bleske-Rechek és Kelsey [10] szerint minél egyszerűbb az előadás, annál magasabb a hallgatók által adott pontszám. A hallgatói értékelések Costin et al. [11] szerint is megbízhatatlanok, mivel a hallgatók inkább az „előadóművészt” díjazták (aki szórakoztat), mint az oktatót. Kutatásaik szerint az értékelések magas korrelációt mutatnak a várt jegyekkel, azaz a köztudottan szigorúan osztályozó oktató rosszabb értékelésre számíthat. Felton et al. [12:106] szerint „a magasabb értékelésnek több köze lesz az oktatók megjelenéséhez vagy ahhoz, hogy mennyire könnyen teljesíthetővé teszik a tárgyat, mint az oktatás minőségéhez”. Ehhez hasonlóan több kutatás is alátámasztotta a tényleges érdemjegyek (az osztályozás észlelt engedékenysége, a tárgy teljesíthetősége) és a hallgatók értékelése közti pozitív kapcsolatot [13], [14], [15], [16], [5]. Felton et al. [17] egészen odáig megy, hogy kijelentse, a hallgatók motivációja az értékelésre az őszinte vágytól az oktató méltó teljesítményének dicséretére egészen az olyan megtorlásig tart, mely nem sokban különbözik a mellékhelyiségekben látható falfirkáktól. Davison és Price [18] szerint manapság a diákokat nem érdekli a tanulási folyamat vagy az eredményeképp létrejövő tudás, a ratemyprofessors.com-on (a továbbiakban RMP) kifejezetten a „megcsömörlött”, jegyorientált, elfogult hallgatók értékelnek. Bleske-Rechek és Kelsey [10] ennél óvatosabban fogalmaz, szerintük ugyanis a

diákok képesek felismerni a jó előadást – ami összhangban van Morgan és Vasché [19] eredményeivel is –, de a megfelelő oktatói minőség észlelése mellett a hallgatók gyakran elfogultak is lehetnek.

Amellett, hogy az értékelések elfogultak lehetnek, azokat bárki (olyan is aki nem látogatta az órákat, esetleg nem is ismeri az oktatót...), bármikor beírhatja (nem csak abban a félévben, amikor tapasztalatait szerezte), illetve nem zárható ki az érzelmi elfogultság, még az is problémát okozhat, hogy az esetek jelentős részében csupán néhány diák küld be értékelést, nem a teljes csoport, így még inkább lehetséges, hogy az eredmények torzítottak [3]. Benton és Cashin [20] kimutatta, hogy az értékelések nagyobb száma oktatónként megbízhatóbbá teszi az átlagos értékelést. Ugyanerre a következtetésre jutott Otto et al. [3] is, szerintük ennek az a magyarázata, hogy ha a hallgatók tapasztalatai nem felelnek meg az oldalon található eredményeknek, akkor erre „kiigazító” értékeléssel válaszolhatnak. Ha például úgy érzik, hogy az oktató túlértékelt, akkor saját tapasztalataikhoz képest is alul, míg ha alulértékeltnek látják, akkor saját valódi véleményükhöz képest felül is pontozhatják az oktatót. Így az átlagértékek közelíthetik a valóságot, megbízhatóbbak lesznek. Ahhoz, hogy megbízhatóak legyenek az oktatói értékelések, Cashin és Perrin [21] szerint legalább 10 értékelés szükséges, ennél kevesebb értékelés csak akkor megbízható, ha azok különböző csoportoktól származnak. Cashin [22] szerint pedig, ha reprezentatív mintát szeretnénk, akkor az évfolyam legalább 2/3-ától kellene adatokat begyűjteni, de ez többnyire nem teljesül az informális oldalak esetén. Ráadásul az értékelések nem feltétlenül egy évfolyamtól, egy csoporttól származnak.

Az informális online értékelést a hallgatók gyakran könnyebbnek érzik, mint a formális értékelést; és mivel névtelenül („arctalanul”) értékelhetnek online, az anonimitást is biztosítottnak látják. Ennek ellenére mégsem minden hallgató értékel, aminek oka Davison és Price [18] szerint az, hogy a hallgatókat megterheli az esetleges regisztráció, a bejelentkezés és maga az értékelés, egyszerűen nem szánják rá az időt. Johnson [23] kutatásai szerint a diákok mindössze 20%-a töltötte ki magától a kutatás céljából összeállított online kérdőívet az oktatóiról. Tanári biztatás hatására ez az arány 32%-ra, kötelező feladatként kiadva 77%-ra, míg plusz pontokért cserébe 87%-ra növekedett. A 100%-ot tehát még így sem lehetett elérni¹.

Akárhogy is, a fentiek miatt az értékelések nem reprezentatívak az összes érintett hallgatóra nézve. A legnagyobb gond, amiben a diákok és az oktatók is egyetértenek, hogy nem tudni, ki értékel az oldalon. Nem tudni, hogy járt-e órára, elment-e az oktató fogadóóráira, elkészítette-e a házi feladatait, tanult-e és végül milyen jegyet kapott [18]. Ez utóbbi problémára az amerikai ratemyprofessors.com (RMP) oldal már megoldást nyújt; ott ugyanis a magyar markmyprofessor.com (a továbbiakban MMP) oldallal ellentétben be lehet írni a hallgató által kapott jegyet és be kell írni a teljesített kurzus kódját. Ez nem csupán a felhasználó ellenőrzésére alkalmas (nagyobb valószínűséggel tudja a tárgy kódját az, aki tényleg teljesítette, felvette), hanem a kiértékelés pontosságát is javítja, hiszen kurzusonként eltérhet az oktató teljesítménye.

A fentiek alapján kutatásunk során két kutatási kérdésre kerestünk választ:

¹ Az értékelés gyakoriságában jelentős eltérések vannak nemek és tudományterületek szerint. Bleske-Rechek és Kelsey [10] vizsgálata szerint a férfi hallgatók nagyobb arányban értékelték, mint a nők (33% vs. 19%), illetve a természettudományi és társadalomtudományi képzésben részt vevők, mint a bölcsészhallgatók.

K1: Milyen összefüggés van az MMP-ben megjelenő, egy adott oktatóhoz tartozó értékelések száma és az adott oktató MMP-ben feltüntetett tulajdonságainak átlagos értékelése között?

Az ehhez kapcsolódó hipotézisünk két dolgon alapul. Egyrészt Davison és Price [18] állításán – mely összhangban van Felton et al. [17] fentebb említett kijelentésével a hallgatók értékelésének motivációjáról –, miszerint az esetleges regisztrációból és értékelésből származó időráfordításnak köszönhetően azok motiváltak az értékelésre, akik vagy nagyon szerették, vagy nagyon nem szerették az oktatóikat. Másrészt pedig Otto et al. [3] szerint a kezdeti értékeket észelve azok, akik ezzel nem értenek egyet, egyfajta kiegyenlítést végeznek. Hipotézisünk ez alapján a következő:

H1: A több hallgatói értékelés rosszabb értékeket eredményez az MMP-ben feltüntetett oktatói tulajdonságokban.

Ez a hipotézis megfelel Otto et al. [3], valamint részben Bleshke-Rechek és Fritsch [24] – RMP adatai alapján megfogalmazott – eredményeinek. Otto et al. [3] úgy vizsgálták az értékelések számának összefüggését az értékelt jellemzőkkel, hogy bevezettek egy bináris változót (egy és több értékelést kapott az oktató), amelynek kapcsolatát vizsgálták az oktató két értékelhető jellemzőjének, az előadásmódjának és a segítőkészségének értékelésével. Regressziós modelljük szerint a több értékelésnek nagyon gyenge, nem szignifikáns negatív hatása van a két értékelhető jellemző átlagértékeire. Bleshke-Rechek és Fritsch [24] pedig nem találtak szignifikáns korrelációt az értékelések száma és az oktatók átlagos minősítése között, viszont nagyon gyenge, de szignifikáns negatív korrelációt találtak az értékelések száma és a kurzus teljesíthetősége között.

K2: Van-e összefüggés az értékelések száma és azok megbízhatósága között a hazai online értékelő felületen?

A fentiekben vázolt szakirodalom ([20], [22], [21], [3]) alapján várakozásaink szerint:

H2: Amennyiben egy oktató több értékeléssel rendelkezik az MMP-ben, az értékelések megbízhatóbbak (ezt mi a kisebb a szórásukkal azonosítjuk).

2. Anyag és módszer

A fenti két hipotézist egy magyar online értékelő felület, a markmyprofessor.com adatain vizsgáltuk. A MMP 2009-ben, az amerikai RMP mintájára jött létre azzal a céllal, hogy segítséget nyújtson a felsőoktatásban tanuló diákoknak és oktató tanároknak. Az oldalon a hallgatók megoszthatják tapasztalataikat az egyes tantárgyakról és oktatókról, illetve az értékelésekből leszűrhetik, hogy mely oktatónál érdemes felvenni egy kurzust, az oktatók pedig visszajelzést kaphatnak a diákoktól [25].

A hallgatók az oldalon akár regisztráció nélkül is értékelhetnek, az alábbi változók mentén és módon:

- tantárgy kiválasztása listából (ha nincs ott, felvihet újat, ellenőrzés nélkül),
- követelmények teljesíthetősége (5 fokozatú skálán, 5 a legjobb),
- tárgy hasznossága (5 fokozatú skálán, 5 a legjobb),
- segítőkészség (5 fokozatú skálán, 5 a legjobb),

- felkészültség (5 fokozatú skálán, 5 a legjobb),
- előadásmód (5 fokozatú skálán, 5 a legjobb),
- „szexisség” (igaz vagy hamis),
- komment (maximum 500 karakteres szöveges értékelés).

A MMP oldalon értékelt oktatók száma folyamatosan nő, a tanulmány írásának időpontjában (2017. március 3.) már 31 296 tanár szerepelt az adatbázisban (összehasonlításképpen az amerikai RMP oldalon 1 685 240 oktató szerepel). Az elemzés alapjául szolgáló adatok gyűjtése 2016. március 15–27. között történt, melynek során az akkor az oldalon szereplő összes, azaz 29 862 db oktatói adatlapot felvettük. Ezen oktatók közül mindössze 20 928 főt (70,08%) értékelték, így a vizsgálatainkban értelemszerűen csak ők szerepeltek. Az oldalon a vizsgálat időpontjában összesen 211 013 értékelés volt. Az oktatók tulajdonságainak és értékelésszámainak átlagértékeit és szórását az 1. táblázat tartalmazza.

Változók	Számtani átlag	Medián	Módusz	Szórás
átlagos értékelés	3,979	4,300	5	1,011
teljesíthetőség	4,015	4,250	5	0,963
hasznosság	3,948	4,110	5	0,963
segítőkézség	3,980	4,330	5	1,090
felkészültség	4,192	4,540	5	0,975
előadásmód	3,772	4,000	5	1,151
értékelések száma	10,380	5,000	1	17,025

1. táblázat: Az értékelt oktatói tulajdonságok és az értékelésszámok átlag- és szórásértékei (N=20928).

(Forrás: Saját számítás a MMP oldalról származó adatok alapján)

Az 1. táblázatban szereplő „átlagos értékelés” a segítőkézség, a felkészültség és az előadásmód egyszerű számtani átlaga; ez az az érték, amely az oktatók adatlapján először szembetűnik a nevük mellett, nagy betűmérettel. Ez alapján az átlag alapján az oktató nevéhez egy angyal (3,66 felett) vagy egy ördög (2,33 alatt) képe kerül [26]. Az átlagos oktatókat nem jelzi az oldal semmilyen speciális módon. A vizsgált mintában 14170 oktató (67,70%) kapott angyal és 1700 (8,12%) ördög minősítést, átlagosnak 5058 oktató (24,17%) bizonyult.

A H1 hipotézis vizsgálatára az MMP oldalról származó, értékeléssel rendelkező oktatók értékelt tulajdonságai és a hozzájuk tartozó értékelésszámok között először Pearson-féle korrelációelemzést alkalmaztunk. Emellett független mintás *t*-próbákkal megvizsgáltuk, hogy a minta értékelésszámok szerinti alsó és felső 10%-ának, illetve az alsó és felső 50%-ának átlagértékei szignifikánsan eltérnek-e. Majd Otto et al. (2008) alapján egy bináris változó segítségével az egy és a több értékeléssel rendelkező oktatók átlagértékeit vetettük össze független mintás *t*-próbákkal. Végül kereszttábla-

elemzést végeztünk az alsó és felső 10, illetve 50%, valamint az egy és több értékeléssel rendelkező oktatók értékelt tulajdonságaira vonatkozóan.

A H2 hipotézis vizsgálatához a minta értékelésszámok szerinti alsó és felső 10%-ának, alsó és felső 50%-ának, valamint az egy és több értékeléssel rendelkező oktatók értékeléseinek szórását vetettük össze Levene-teszt segítségével, ugyanis az értékeléseket annál megbízhatóbbnak mondhatjuk, minél kisebb azok szórása.

3. Eredmények

Az első hipotézis (H1) szerint a több hallgatói értékelés rosszabb értékeket eredményez az értékelt oktatói tulajdonságokban. Ennek vizsgálatához először az átlagos értékelés, a teljesíthetőség, a hasznosság, a segítőkészség, a felkészültség és az előadásmód értékeinek korrelációját vizsgáltuk az értékelések számával. Az eredményeket a 2. táblázat tartalmazza.

Változók	lineáris korreláció az értékelések számával
átlagos értékelés	-0,054***
teljesíthetőség	-0,108***
hasznosság	-0,052***
segítőkészség	-0,079***
felkészültség	-0,042***
előadásmód	-0,033***

2. táblázat: Az értékelések száma és az egyes értékelt tulajdonságok értéke közötti Pearson-féle korreláció (N = 20928).

Megjegyzés: *** 1%-en szignifikáns

(Forrás: Saját számítás a MMP oldalról származó adatok alapján)

Ahogy azt a 2. táblázat mutatja, az értékelések száma valamennyi tulajdonsággal gyenge negatív korrelációt mutat. Azaz az értékelések számának növekedésével nagyon enyhe valószínűséggel romlik az oktatók átlagos megítélése az egyes tulajdonságaikban (ez természetesen ellentétes irányú kapcsolatra is utalhat: a rosszabb megítélésű oktatókat nagyon kicsit hajlamosabbak gyakrabban értékelni a hallgatók). Bár a korrelációk 1%-on szignifikánsak, nagyon gyengék (praktikusan nézve korrelátlanságról beszélhetünk), ezért érdemes tovább vizsgálni az adatokat.

Még pontosabb képet kaphatunk az összefüggésekről akkor, ha az értékelések száma szerint sorba rendezett minta alsó és felső valahány százalékának tulajdonság-értékeit is összevetjük egymással. Ezért független mintás *t*-próba segítségével megvizsgáltuk, hogy a minta alsó és felső 10%-ának, illetve az alsó és felső 50%-ának értékei szignifikánsan eltérnek-e. Az eredményeket a 3. és 4. táblázat tartalmazza. Az eredmények megerősítik az előző korrelációs vizsgálatokat, azaz az alsó 10% és 50%

szignifikánsan magasabb átlagos értékeket kapott minden tulajdonság esetén, mint a felső 10% és 50%. Tehát az értékelések számának növekedésével romlik az oktatók átlagos megítélése.

Az Otto et al. [3] által bevezetett bináris változót alkalmazva mi is megvizsgáltuk, hogy az értékelések átlaga mutat-e eltérést aszerint, hogy az oktató egy vagy több értékelést kapott-e.² A független mintás *t*-próbák eredményeit az 5. táblázat mutatja. Ebből az derül ki, mint a korábbi vizsgálatokból: a több értékelés alacsonyabb átlagos értékekkel jár együtt az összes jellemző esetén.

A fentiek alapján tehát kijelenthetjük, hogy a több értékelés valóban alacsonyabb átlagokkal jár együtt az összes értékelt jellemző tekintetében, így a H1 hipotézist elfogadjuk.

Változók	alsó / felső 10%	átlag	szórás	<i>t</i> -érték	<i>F</i> -érték
átlag	alsó	4,176	1,214	83,837***	11,392***
	felső	3,796	0,923		
teljesíthetőség	alsó	4,222	1,156	83,193***	16,380***
	felső	3,697	0,901		
hasznosság	alsó	4,083	1,275	408,805***	9,457***
	felső	3,773	0,786		
segítőkézség	alsó	4,225	1,309	87,683***	14,244***
	felső	3,713	0,994		
felkészültség	alsó	4,337	1,225	152,629***	8,790***
	felső	4,051	0,847		
előadásmód	alsó	3,970	1,419	211,126***	8,828***
	felső	3,633	1,016		

3. táblázat: A minta alsó és felső 10%-ának vizsgálata (*N* = 4184).

Megjegyzés: *** 1%-on szignifikáns; a *t*-érték a kétmintás *t*-teszt, az *F*-érték a Levene-teszt statisztikája

(Forrás: saját számítás a MMP oldalról származó adatok alapján)

Változók	alsó / felső 50%	átlag	szórás	<i>t</i> -érték	<i>F</i> -érték
átlag	alsó	4,066	1,094	289,835***	12,499***
	felső	3,892	0,912		
teljesíthetőség	alsó	4,143	1,026	188,333***	19,502***
	felső	3,886	0,877		

² Az értékeléssel rendelkező oktatók teljes mintájából 3594 oktatonál (17,17%) található egy, 17334 oktatonál (82,83%) pedig több értékelés.

hasznosság	alsó	4,027	1,089	780,371***	11,982***
	felső	3,868	0,810		
segítőkézség	alsó	4,098	1,179	299,610***	15,839***
	felső	3,861	0,980		
felkészültség	alsó	4,254	1,080	440,827***	9,276***
	felső	4,129	0,853		
előadásmód	alsó	3,851	1,262	541,152***	9,907***
	felső	3,694	1,023		

4. táblázat: A minta alsó és felső 50%-ának vizsgálata (N = 20928).

Megjegyzés: *** 1%-en szignifikáns; a t-érték a kétmintás t-teszt, az F-érték a Levene-teszt statisztikája

(Forrás: saját számítás a MMP oldalról származó adatok alapján)

Változók	alsó / felső 50%	átlag	szórás	t-érték	F-érték
átlag	egy	4,165	1,229	294,238***	10,347***
	több	3,940	0,955		
teljesíthetőség	egy	4,211	1,171	371,616***	11,415***
	több	3,974	0,909		
hasznosság	egy	4,083	1,279	916,375***	7,300***
	több	3,920	0,881		
segítőkézség	egy	4,209	1,330	334,837***	11,789***
	több	3,932	1,027		
felkészültség	egy	4,331	1,231	388,274***	7,732***
	több	4,163	0,911		
előadásmód	egy	3,960	1,432	584,876***	8,969***
	több	3,734	1,080		

5. táblázat: Az értékelések számának kapcsolata a tulajdonságokkal, egy és több értékelés esetén (N = 20928).

Megjegyzés: *** 1%-en szignifikáns; a t-érték a kétmintás t-teszt, az F-érték a Levene-teszt statisztikája

(Forrás: saját számítás a MMP oldalról származó adatok alapján)

Érdekes lehet azonban közelebbről is megvizsgálni, hogy mi lehet az oka a negatív kapcsolatnak, ugyanis a korrelációk tulajdonképpen elfedik a köztes ingadozásokat. Ezért keresztábra-elemzéseket is végeztünk, az egy és több értékelést mutató bináris változó és az értékelt jellemzők egész számokra

kerekített átlagértékeinek összevetésével.³ Az elemzés szignifikáns eltéréseket mutatott ki valamennyi jellemző tekintetében (átlagos értékelés: $\chi^2=1247,257$; teljesíthetőség: $\chi^2=1062,097$; hasznosság: $\chi^2=1579,427$; segítőkézség: $\chi^2=1578,458$; felkészültség: $\chi^2=1221,601$; előadásmód: $\chi^2=1516,352$; $p<0,001$ minden esetben). Az összes jellemző esetén szinte ugyanolyan keresztábrákat kaptunk, ezért itt csak egyet (a segítőkézségre vonatkozót) mutatunk be példaképpen (ld. 6. táblázat és 1. ábra).

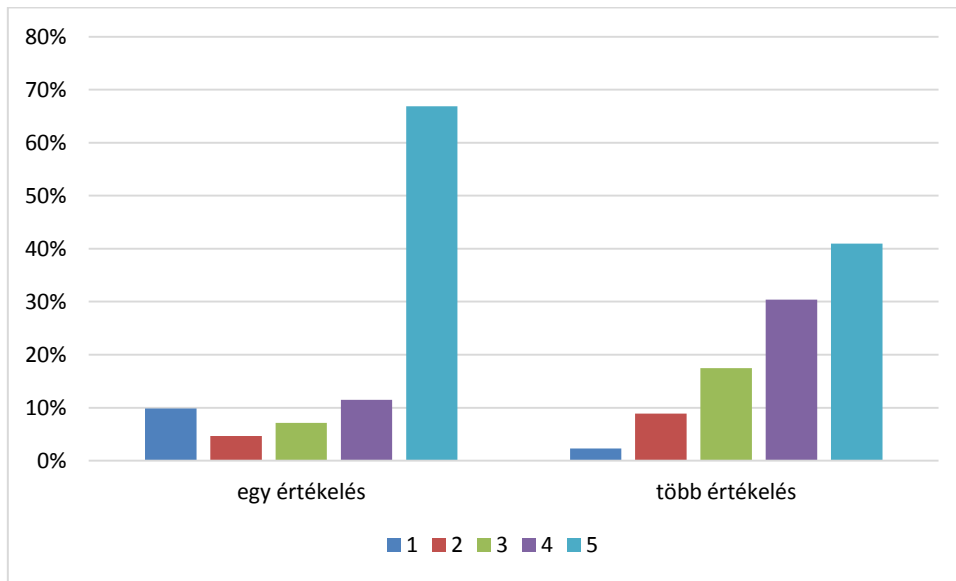
A 6. táblázatból és az 1. ábrából világosan látszik, hogy az egyes és ötös értékelések az egy értékeléssel rendelkező oktatóknál felül-, míg a több értékeléssel rendelkezőknél alulreprezentáltak; a kettes, hármas, négyes értékelések pedig pont fordítva, az egy értékeléssel rendelkezőknél alul-, a több értékeléssel rendelkezőknél felülreprezentáltak. Ez azt jelenti, hogy az először értékelő hallgatók sokkal inkább adnak szélsőséges értékeket (egyest vagy ötöst), ami pedig arra utal, hogy először azok adnak értékelést, akik vagy nagyon elégedettek, vagy nagyon elégedetlenek. A korábban ismertetett nagyon gyenge negatív korrelációt az okozza, hogy a nagy számban előforduló ötösök aránya az értékelések növekedésével jelentősen csökkent (az induló arány 41,23%-ával), míg az egyesek arányának csökkenése (78,85%-kal) kisebb hatással volt az átlagra a kisebb számarányuk miatt, hiába volt arányuk csökkenése közel kétszer akkora, mint az ötös értékelések arányáé.

		1	2	3	4	5	Össz.
egy értékelés	Darab	195	78	127	228	1256	1884
	értékelések száma szerinti megoszlás (%)	10,4%	4,1%	6,7%	12,1%	66,7%	100,0%
	Korrigált stand. reziduum	17,5	-7,5	-12,3	-16,7	22,0	
több értékelés	Darab	212	900	1747	2974	3755	9588
	értékelések száma szerinti megoszlás (%)	2,2%	9,4%	18,2%	31,0%	39,2%	100,0%
	Korrigált stand. reziduum	-17,5	7,5	12,3	16,7	-22,0	
Összesen	Darab	407	978	1874	3202	5011	11472
	értékelések száma szerinti megoszlás (%)	3,5%	8,5%	16,3%	27,9%	43,7%	100,0%

6. táblázat: Az értékelések száma és a segítőkézség keresztábrája (N = 20928).

(Forrás: saját számítás a MMP oldalról származó adatok alapján)

³ A keresztábrás vizsgálatokat elvégeztük a minta alsó és felső 10, illetve 50%-ára vonatkozóan is, és a fentivel teljes mértékben megegyező mintázatot (természetesen eltérő számértékekkel) találtunk minden jellemző tekintetében.



1. ábra: Az értékelések száma és a segítőkészség kereszt táblája (N = 20928).

(Forrás: saját számítás a MMP oldalról származó adatok alapján)

A H2 hipotézist az értékelések száma szerint sorba rendezett minta alsó és felső 10, illetve 50%-a, valamint az egy és több értékeléssel rendelkezők értékeléseinek szórása közti különbségek összevetésével vizsgáltuk. A szórások különbsége minden értékelt jellemző esetében szignifikánsnak bizonyult a Levene-teszt alapján (lásd 3., 4., 5. táblázat), így kijelenthetjük, hogy a több értékelés a szórás csökkenésével jár, azaz megbízhatóbb lesz. Tehát a H2 hipotézist a szakirodalmi megállapításokkal összhangban elfogadjuk.

4. Konklúzió

Kutatásunkban arra kerestük a választ, hogy az oktatók hallgatói értékelésére szolgáló online oldalakon szereplő értékelések megbízhatósága függ-e az értékelések számától, illetve az értékelések átlaga mutat-e összefüggést az értékelések számával. A markmyprofessor.com oldalról származó adatok alapján megállapítottuk, hogy az értékelések száma szinte elenyészően gyenge, de szignifikáns negatív lineáris kapcsolatot mutat az értékek átlagával. Jelentősebb mértékű, ugyanilyen előjelű eltérést találtunk a leggyakrabban értékelték és a legritkábban értékelték csoportjainak átlagos „osztályzatai” között is. Azaz minél több értékelést kap egy oktató, annál alacsonyabbak az értékelt tulajdonságok átlagai.

Az értékelések számának növekedésével emellett az értékek szórása csökken, ami az adatok megbízhatóságának (reliabilitásának) növekedésére utal, bár azok érvényességét (validitását) nem feltétlenül érinti. A kevés értékelést kapott oktatókat tehát inkább magasabbra értékelik, majd ahogyan nő az értékelések száma, a jellemzők átlagértékei egy alacsonyabb érték felé tartanak, és egyben egyre hasonlóbakká válnak. Továbbá az is állítható, hogy az első értékelést adó diákok hajlamosabbak szélsőséges értékeléseket (egyesekeket vagy ötösöket) adni.

A fentiekből következik, hogy egy online értékelő oldalakon kevés értékeléssel rendelkező oktatót nem jellemez megbízhatóan az átlagos értékelés, így abból nem lehet érdemi következtetéseket levonni. Minél több értékeléssel rendelkezik viszont egy oktató, annál megbízhatóbban ítéhető meg az online felületen véleményezőik által „megítélt” teljesítménye. Fontos ehhez hozzátenni, hogy adataink nem voltak alkalmasak a válaszok érvényességének tesztelésére, azaz azt nem tudtuk vizsgálni, vajon a több értékelés jelent-e elmozdulást a „valós” értékelés felé (azaz ami az összes tanított hallgató véleményét reprezentálná).

További kutatási irány lehet egyrészt annak a vizsgálata, hogy az értékelések hogyan függnek egyéb, az oktatót vagy a kurzust jellemző változóktól, például az oktató nemétől, korától, kinézetétől, az oktatott tárgy jellegétől, a tudományterülettől, másrészt hogy az értékelt jellemzők mutatnak-e összefüggést egymással, és ha igen, az milyen irányú.

Hivatkozások

- [1] E. L. Tipoe (2013), *Course Enrollment Decisions: The Substitutability of Online Ratings Websites for Official Professor Evaluations*, LAP Lambert Academic Publishing, Saarbrücken
- [2] W. E. Cashin (1996), *Developing an Effective Faculty Evaluation System*, IDEA Paper No. 33, http://ideaedu.org/wp-content/uploads/2014/11/Idea_Paper_33.pdf, letöltés ideje: 2015. 09. 15.
- [3] J. Otto, D. A. Sanford, D. N. Ross (2008), *Does ratemyprofessor.com really rate my professor?* Assessment & Evaluation in Higher Education, 33. évf. 4. sz. 355-368.
- [4] T. Coladarci, I. Kornfield (2007), *RateMyProfessors.com versus formal in-class student evaluations of teaching*, Practical Assessment, Research & Evaluation, 12. évf. 6. sz. 1-15.
- [5] J. J. Siegfried, R. Fels (1979), *Research on teaching college economics: A survey*, Journal of Economic Literature, 17. évf. 3. sz. 923-969.
- [6] W. E. Cashin (1989), *Defining and Evaluating College Teaching*, IDEA Paper No. 21, http://ideaedu.org/wp-content/uploads/2014/11/Idea_Paper_21.pdf, letöltés ideje: 2015. 09. 14.
- [7] A. G. Greenwald, G. M. Gillmore (1997), *Grading leniency is a removable contaminant of student ratings*, American Psychologist, 52. évf. 11. sz. 1209-1217.
- [8] R. Wilson (1998), *New research casts doubt on value of student evaluations of professors*, Chronicle of Higher Education, 44. évf. 19. sz. A12-A14.
- [9] S. Liaw, K. Goh (2003), *Evidence and control of biases in student evaluations of teaching*, International Journal of Educational Management, 17. évf. 1. sz. 37-43.
- [10] A. Bleske-Rechek, M. Kelsey (2010), *RateMyProfessors.com: Testing Assumptions about Student Use and Misuse*, Practical Assessment, Research & Evaluation, 15. évf. 5. sz. 1-12.

- [11] F. Costin, W. T. Greenough, R. J. Menges (1971), *Student Ratings of College Teaching: Reliability, Validity, and Usefulness*, Review of Educational Research, 41. évf. 5. sz. 511-535.
- [12] J. Felton, J. Mitchell, M. Stinson (2004), *Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness*, Assessment & Evaluation in Higher Education. 29. évf. 1. sz. 91-108.
- [13] D. R. Capozza (1973), *Student Evaluations, Grades and Learning in Economics*, Western Economic Journal, 11. évf. 1. sz. 127.
- [14] J. C. Soper (1973), *Soft Research on a Hard Subject: Student Evaluations Reconsidered*, Journal of Economic Education, 5. évf. 1. sz. 22-26.
- [15] R. B. McKenzie (1975), *The Economic Effects of Grade Inflation on Instructor Evaluations: A Theoretical Approach*, Journal of Economic Education, 6. évf. 2. sz. 99-106.
- [16] J. B. Kau, P. H. Rubin (1976), *Measurement Techniques, Grades and Ratings of Instructors*. Journal of Economic Education, 8. évf. 1. sz. 59-62.
- [17] J. Felton, P. Koper, J. Mitchell, M. Stinson (2008), *Attractiveness, easiness, and other*, Assessment & Evaluation in Higher Education, 33. évf. 1. sz. 45-61.
- [18] E. Davison, J. Price (2009), *How do we rate? An evaluation of online student evaluations*, Assessment & Evaluation in Higher Education, 34. évf. 1. sz. 51-65.
- [19] W. D. Morgan, J. D. Vaché (1978), *An Educational Production Function Approach to Teaching Effectiveness and Evaluation*, Journal of Economic Education, 9. évf. 2. sz. 123-26.
- [20] S. L. Benton, W. E. Cashin (2009), *Student ratings of teaching: the research revisited*, IDEA Paper No. 50, http://ideaedu.org/wp-content/uploads/2014/11/idea-paper_50.pdf, letöltés ideje: 2015. 10. 01.
- [21] W. E. Cashin, B. Perrin (1978), *Description of IDEA Standard Form data base*, IDEA technical report no. 4, Kansas State University, Center for Faculty Evaluation and Development, Manhattan.
- [22] W. E. Cashin (1990), *Students do rate different academic fields differently*, New Directions for Teaching and Learning, 43. évf. 113-121.
- [23] T. Johnson (2002), *Online Student Ratings: Will Student Respond?* Annual Conference of the American Educational Research Association, New Orleans, April 1-5, 2002.
- [24] A. Bleshke-Rechek, A. Fritsch (2011), *Student Consensus on RateMyProfessors.com*, Practical Assessment, Research & Evaluation, 16. évf. 18. sz.
- [25] MMP (2015a): Rólunk. <http://www.markmyprofessor.com/rolunk.html>, letöltés ideje: 2015. 09. 10.
- [26] MMP (2015b): Módszertan. <http://www.markmyprofessor.com/modszertan.html>, letöltés ideje: 2015. 09. 10.